DAVID DEUTSCH

LA ESTRUCTURA DE LA REALIDAD

Título original: *The Fabric of the Reality*

David Deutsch, Editorial Anagrama, S. A., 1999.

Traducción: David Sempau

Ilustraciones: Pep Montserrat

Diseño/retoque portada: Julio Vivas

AGRADECIMIENTOS

El desarrollo de las ideas expuestas en este libro se vio considerablemente favorecido por conversaciones con Brice DeWitt, Artur Ekert, Michael Lockwood, Enrico Rodrigo, Dennis Sciama, Frank Tipler, John Wheeler y Kolya Wolf.

Doy las gracias a mis amigos y colegas Ruth Chang, Artur Ekert, David Johnson-Davies, Michael Lockwood, Enrico Rodrigo y Kolya Wolf, a mi madre, Tikvah Deutsch, así como a mis editores Caroline Knight y Ravi Mirchandani (de Penguin Books), John Woodruff y, especialmente, Sarah Lawrence, por su exhaustiva y crítica lectura de los primeros borradores y sus sugerencias de múltiples correcciones y mejoras.

También estoy agradecido a Harvey Brown, Steve Graham, Rossella Lupaccini, Svein Olav Nyberg, Oliver y Harriet Strimpel y, especialmente, Richard Dawkins y Frank Tipler, por haber leído y comentado diversos fragmentos del original.

PREFACIO

El principal motivo que me ha inducido a exponer la concepción del mundo que se ofrece en este libro es que, gracias a una serie de extraordinarios descubrimientos científicos, disponemos hoy día de algunas teorías muy complejas sobre la estructura de la realidad. Si aspiramos a comprender el mundo en un sentido que no sea meramente superficial, deberemos hacerlo por medio de esas teorías y la razón, y no basándonos en nuestros prejuicios, las ideas que nos han sido imbuidas o, incluso, el sentido común. Las mejores de esas teorías no sólo son más ciertas que el sentido común, sino que también son mucho más sensatas. Debemos tomárnoslas en serio, no como meros fundamentos pragmáticos para sus respectivos campos, sino como explicaciones del mundo. Opino también que alcanzaremos la máxima comprensión de la realidad si las consideramos de manera conjunta y no por separado, puesto que están inextricablemente relacionadas.

Quizá parezca extraño que la sugerencia de que debemos tratar de comprender el mundo de modo racional y coherente, y para ello hemos de recurrir a las mejores y más fundamentales de esas teorías, pueda resultar insólita o suscitar controversias, pero así es. Una de las razones de esta situación es que dichas teorías, cuando se toman en serio, presentan múltiples implicaciones que no se pueden comprender de modo intuitivo. Consecuencia de ello es que se ha hecho toda clase de intentos para evitar enfrentarse a esas implicaciones, desde introducir modificaciones ad hoc en dichas teorías o reinterpretarlas, hasta reducir de manera arbitraria su ámbito de aplicación o, simplemente, utilizarlas en la práctica, pero sin aspirar a sacar de ellas todas las conclusiones posibles. Criticaré algunos de estos intentos (ninguno de los cuales, en mi opinión, tiene demasiado mérito), pero sólo cuando ello contribuya a explicar mejos las propias teorías. Y es que la presente obra no tiene como objetivo prioritario defender dichas teorías, sino investigar cuál sería la estructura más verosimil de la realidad si fueran ciertas.

1. LA TEORÍA TOTAL DE LA REALIDAD

Recuerdo que me contaban, siendo niño, que en la antigüedad era posible que una persona aprendiese la totalidad del saber de su época. Me decían también que hoy día son tantos los conocimientos, que nadie puede aspirar a aprender más que una minúscula fracción, incluso a lo largo de una dilatada vida. Ésta última afirmación me sorprendía y me contrariaba. De hecho, me negaba a creerla. No sabía cómo justificar mi incredulidad, pero lo cierto es que no quería que las cosas fuesen así, y envidiaba a los sabios de la antigüedad.

No era que pretendiese memorizar todos los hechos recogidos en la infinidad de las enciclopedias que hay en el mundo; bien al contrario, me repugnaba la memorización. No era éste el sentido en que deseaba aprender todo el saber de mi época. No me habría desanimado que me hubiesen dicho que aparecen diariamente más publicaciones de las que podría leer en toda mi vida, o que existen seiscientas mil especies conocidas sólo de escarabajos. No albergaba ningún deseo de seguir la trayectoria de cada alondra. No imaginaba tampoco que un sabio de la antigüedad que, hipotéticamente hubiese aprendido todos los conocimientos de su tiempo, lo hubiese hecho así. En realidad, tenía una idea más sutil de lo que debía considerarse aprender: por «aprender» entendía *comprender*.

La idea de que una persona sea capaz de comprender todo lo comprensible puede parecer no menos fantástica, pero, ciertamente, no lo es tanto como la de que una persona memorice todos los hechos conocidos. Por ejemplo, nadie podría memorizar todos los datos publicados como consecuencia de las observaciones ni siquiera en un campo tan reducido como es el de los movimientos de los planetas, pero son muchos los astrónomos que *comprenden* esos movimientos en la medida en que es posible comprenderlos. Y ello es así porque la comprensión no depende del conocimiento de una infinidad de hechos aislados, sino de disponer de los conceptos, explicaciones y teorías adecuados. Una teoría relativamente sencilla y comprensible puede englobar un número infinito de hechos de difícil asimilación. La mejor teoría sobre el movimiento de los planetas es *la teoría general de la relatividad* de Einstein, que a principios del siglo XX dejó obsoletas

la ley de la gravitación universal y las demás leyes de Newton, fundamento de la mecánica clásica. En principio, la teoría general de la relatividad es capaz de predecir correctamente no sólo *todos* los movimientos planetarios, sino también *todos* los demás efectos de la gravitación con la máxima exactitud que permiten nuestros sistemas de medidas. Que una teoría prediga algo «en principio» significa que las predicciones se siguen lógicamente de ella, incluso si en la práctica los cálculos necesarios a fin de realizar algunas de esas predicciones son demasiado complejos para resultar factibles desde un punto de vista tecnológico, o incluso para que sea posible desarrollarlos en el universo tal como lo conocemos.

La capacidad de predecir o describir las cosas, aun con la máxima exactitud, no significa, ni mucho menos, comprenderlas. En física las predicciones y las descripciones se expresan a menudo mediante fórmulas matemáticas. Supongamos que, si dispusiese del tiempo y la inclinación para ello, memorizara la fórmula que me permitiría calcular cualquier posición planetaria recogida en las tablas astronómicas. ¿Qué ganaría con ello, exactamente, en comparación con memorizar directamente dichas tablas? La fórmula es más fácil de recordar, pero, por otra parte, buscar un dato en las tablas puede ser más fácil todavía que calcularlo mediante la fórmula. La verdadera ventaja que ofrece ésta es que puede ser utilizada incluso en casos que no están recogidos en las tablas, por ejemplo, para predecir el resultado de futuras observaciones. Puede también situar más exactamente las posiciones históricas de los planetas, pues los datos recogidos en las tablas contienen errores de observación. Con todo, aunque la fórmula permite el acceso a un número de datos mucho mayor que el recogido en las tablas, sabérsela de memoria no equivale a comprender el movimiento de los planetas. Los hechos no pueden ser entendidos simplemente porque sean expresados mediante una fórmula, al igual que no lo pueden ser por estar recogidos en unas tablas o ser confiados a la memoria. Sólo pueden ser comprendidos si son explicados. Afortunadamente, las principales teorías ofrecen tanto profundas explicaciones como ajustadas predicciones. La teoría general de la relatividad, por ejemplo, explica la gravedad en términos de una nueva geometría tetradimensional de tiempo y espacio curvos. Explica con toda precisión cómo esta geometría afecta a la materia y es afectada por ella. Esta explicación constituye el contenido fundamental de la

teoría; las predicciones sobre movimientos planetarios no son más que algunas de las consecuencias que podemos deducir de esa explicación.

Lo que hace tan importante a la teoría general de la relatividad no es que permita predecir los movlmientos planetarios un poco más exactamente que la teoría de Newton, sino el hecho de revelar y explicar aspectos de la realidad previamente insospechados, tales como la curvatura del espacio y el tiempo. Esto es algo característico de la explicación científica. Las teorías científicas explican los objetos y fenómenos que conocemos por medio de la experiencia en términos de una realidad subyacente que no experimentamos de modo directo. Pero la capacidad de una teoría para explicar lo que conocemos sólo por la experiencia no es su más valioso atributo, sino el hecho de que explique la estructura de la realidad. Como veremos, uno de los atributos más valiosos y significativos, así como útiles, del pensamiento humano, en general, es su capacidad para descubrir y explicar la estructura de la realidad.

Algunos filósofos, sin embargo —e incluso algunos científicos—, menosprecian el papel de la explicación científica. Para ellos, el propósito básico de una teoría científica no es explicar todo, sino predecir los resultados de la experimentación: su contenido fundamental reside en la capacidad de predicción de sus fórmulas. Consideran que cualquier explicación consistente que una teoría pueda ofrecer para sus predicciones resulta válida —o, incluso, que tanto da que no ofrezca ninguna explicación—, mientras esas predicciones sean ciertas. Este punto de vista recibe el nombre de *instrumentalismo* ya que considera que la teoría es un mero «instrumento» para hacer predicciones. Para los instrumentalistas, la idea de que la ciencia pueda permitirnos comprender la realidad subyacente que es la causa última de los fenómenos que descubren nuestras observaciones es una falacia y una vana presunción. Opinan que las teorías sólo son válidas para predecir los resultados de los experimentos, y todo lo demás es mera palabrería. Las explicaciones, en particular, son consideradas por ellos simples apoyos psicológicos: una especie de ficción que incorporamos a las teorías para hacerlas más entretenidas y fáciles de recordar. El premio Nobel de Física Steven Weinberg se encontraba, sin duda, en un estado de ánimo instrumentalista

cuando formuló este insólito comentario acerca de la explicación que da Einstein de la gravedad:

Lo importante es que podamos hacer predicciones acerca de las imágenes que aparecen en las placas fotográficas de los astrónomos o las frecuencias de las rayas espectrales, por ejemplo; en cambio, carece de importancia que atribuyamos esas predicciones a los efectos físicos de los campos gravitatorios sobre el movimiento de los planetas y los fotones [como en la física anterior a Einstein] o a una curvatura del espacio y el tiempo (*Gravitation and Cosmology*, p. 147).

Weinberg y los demás instrumentalistas se equivocan. De hecho, *importa* a qué atribuimos las imágenes que aparecen en las placas de los astrónomos, e importa no sólo a los físicos teóricos, como yo, cuya motivación básica para la formulación y el estudio de teorías es el deseo de comprender mejor el mundo. Y, sin duda, ésa es también la motivación de Weinberg. ¡No creo que sienta realmente la compulsión de predecir imágenes y espectros! Y es que, incluso en aspectos meramente prácticos, la capacidad explicativa de una teoría es lo importante, y la predictiva lo accesorio. Si esto resulta sorprendente, imaginemos que un científico extraterrestre hubiese visitado la Tierra y nos hubiese dejado un «oráculo» de tecnología ultraalta que permitiese predecir el resultado de cualquier experimento, pero no ofreciese explicación alguna del porqué. Según los instrumentalistas, una vez en posesión de semejante oráculo, ya no necesitaríamos teoría científica alguna, excepto, quizás, como entretenimiento. Pero ¿es esto cierto? ¿Cómo se utilizaría el oráculo en la práctica? Es evidente que debería poseer el conocimiento para construir una nave espacial, por ejemplo. Pero ¿cómo nos ayudaría exactamente a construirla, o a construir otro oráculo idéntico a él, o incluso una ratonera más eficaz? El oráculo sólo predeciría los resultados de los experimentos; por lo tanto, antes de consultarle, deberíamos conocer los experimentos sobre los que le íbamos a interrogar. Si le suministrásemos el diseño de la nave espacial y los detalles de una determinada prueba de vuelo, nos diría cómo se comportaría mientras era probada, pero en ningún caso diseñaría la nave por

nosotros. Incluso si predijera que iba a estallar en el despegue, no nos diría cómo prevenir la explosión. Esta tarea también quedaría reservada para nosotros, pero antes de emprenderla, antes incluso de intentar mejorar en lo posible su diseño, deberíamos haber *comprendido*, entre otras cosas, el funcionamiento de la nave. Sólo entonces tendríamos la posibilidad de descubrir la existencia de algún fallo capaz de hacerla estallar durante el despegue. La predicción —incluso la universal, perfecta—, simplemente, no sustituye a la explicación.

Del mismo modo, en una investigación científica el oráculo no nos proporcionaría ninguna teoría nueva. A menos que ya dispusiésemos de alguna y se nos hubiese ocurrido el experimento para probarla, no podríamos preguntarle al oráculo qué sucedería si esa teoría era sometida a dicha prueba. Por lo tanto, el oráculo no sustituiría a las teorías, sino a los experimentos. Nos ahorraría cuantiosas inversiones en laboratorios y aceleradores de partículas. En lugar de construir prototipos de naves espaciales y arriesgar la vida de pilotos de pruebas, podríamos realizar todos los ensayos en tierra, con pilotos cómodamente instalados en simuladores de vuelo cuyo comportamiento sería controlado por las predicciones del oráculo.

El oráculo sería útil en múltiples situaciones, pero su utilidad dependería siempre de nuestra capacidad para resolver los problemas científicos del mismo modo en que lo hacemos ahora, es decir, ideando teorías explicativas. Ni siquiera reemplazaría a *toda* la experimentación, puesto que su utilidad para predecir los resultados de un determinado experimento dependería, en la práctica, de si describírselo de la manera adecuada para que proporcionase una respuesta útil fuera más fácil que realizarlo directamente. Es evidente que el oráculo necesitaría una especie de «interfaz de usuario». Lo más probable es que hubiera que introducir en él una descripción lo más completa posible del experimento mediante algún lenguaje estándar, lenguaje en el que ciertos experimentos resultarían más difíciles de describir que otros. En la práctica, la descripción necesaria para muchos de los experimentos resultaría demasiado compleja para poderla introducir. El oráculo presentaría, pues, las mismas ventajas e inconvenientes que las otras fuentes de datos experimentales, y sólo resultaría útil en los casos en que su empleo fuese más conveniente que el de cualquiera de ellas. En realidad, ya

disponemos de dicho oráculo: es el mundo físico que nos rodea. Nos dice el resultado de cualquier experimento concebible siempre que se lo preguntemos utilizando el lenguaje adecuado (por ejemplo, realizándolo), si bien en algunos casos no resulta práctico «introducir la descripción del experimento» en la forma requerida (por ejemplo, construir el aparato y hacerlo funcionar), pero no nos ofrece explicaciones.

En contadas aplicaciones, por ejemplo, en la predicción meteorológica, un oráculo puramente predictivo puede resultarnos casi tan satisfactorio como una teoría explicativa. Pero ello sólo sería así en el caso de que las predicciones del oráculo fuesen completas y perfectas. En la práctica, como bien sabemos, las predicciones meteorológicas son incompletas e imperfectas, y, para compensarlo, incluyen explicaciones de cómo los meteorólogos llegaron a sus conclusiones. Estas explicaciones nos permiten juzgar la fiabilidad de las predicciones y deducir otros pronósticos relevantes para el lugar donde nos encontramos y nuestras necesidades. Por ejemplo, es importante saber si la predicción de que mañana será un día ventoso se debe a que va a afectarnos una zona de altas presiones o a que se acerca un huracán: en este último caso, tomaremos las necesarias precauciones. Los meteorólogos necesitan también teorías explicativas sobre el tiempo, a fin de saber qué variables deben introducir en sus programas informáticos de predicción, qué observaciones adicionales será oportuno agregar para conseguir un pronóstico más ajustado, etcétera.

Así pues el ideal instrumentalista encarnado por nuestro imaginario oráculo —es decir, una teoría científica a la que se ha despojado de todo su contenido explicativo— resultaría de una utilidad estrictamente limitada. Debemos estar satisfechos de que las teorías científicas reales no se parezcan a este ideal y de que, en la realidad, los científicos no dirijan sus esfuerzos a conseguirlo.

Una forma extrema de instrumentalismo, llamada *positivismo*, *neopositivismo* o *positivismo lógico*, sostiene que todo enunciado que no describa o prediga observaciones, no sólo es superfluo, sino que carece de sentido. Si bien esta doctrina carece asimismo de todo sentido, según sus propios criterios, constituyó la teoría dominante en el conocimiento científico de la primera mitad del siglo XX, y las ideas instrumentalistas y positivistas aún

tienen seguidores. Una de las razones por las que resultan superficialmente plausibles es el hecho de que, aunque la predicción no es el propósito de la ciencia, forma parte de su *método* característico de operar. El método científico implica que cada vez que se postula una nueva teoría para explicar un fenómeno, debe llevarse a cabo una *prueba experimental crucial*, un experimento que compara las predicciones de la vieja teoría y la nueva. Ello permite descartar aquella cuya predicción resultó errónea. Así pues, el resultado de la prueba experimental crucial para decidir entre dos teorías depende de las predicciones de éstas, y no de sus explicaciones. Éste es el origen de la idea errónea de que lo único que cuenta en una teoría científica son sus predicciones. Pero la prueba experimental no es, ni mucho menos, el único proceso involucrado en el progreso del conocimiento científico. La gran mayoría de las teorías se rechazan porque contienen explicaciones defectuosas, no porque no superen las pruebas experimentales. Las rechazamos incluso sin molestarnos en probarlas. Imaginemos, por ejemplo, la teoría de que comer un kilo de hierba constituye una cura efectiva contra el resfriado común. Esta teoría propone predicciones experimentalmente comprobables: si fuese ensayada con resultados nulos, sería rechazada como falsa. Pero nunca ha sido ensayada—ni lo será, con toda probabilidad—, porque no ofrece ninguna explicación de qué causa la cura. Con toda razón, la presumimos falsa. Continuamente aparecen infinidad de nuevas teorías posibles de esa clase, compatibles con observaciones existentes y que ofrecen nuevas predicciones, de modo que jamás dispondremos del tiempo y los medios necesarios para comprobarlas todas. Lo que sometemos a prueba son aquellas nuevas teorías que parecen ofrecer la posibilidad de proporcionar mejores explicaciones que las que las han precedido.

Afirmar que la predicción es el propósito de una teoría científica es confundir los medios con los fines. Es como decir que el propósito de una nave espacial es quemar combustible. De hecho, consumir combustible es sólo una de las muchas cosas que una nave espacial debe hacer para alcanzar su verdadera finalidad, que es transportar su carga de un punto del espacio a otro. Superar pruebas experimentales no es más que una de las muchas cosas que una teoría debe hacer para alcanzar la verdadera finalidad de la ciencia, que es explicar el mundo.

Como he dicho, es inevitable que del entramado de las explicaciones formen parte cosas que no observamos directamente: átomos y fuerzas, el interior de las estrellas y la rotación de las galaxias, el pasado y el futuro, las leyes de la naturaleza. Cuanto más profunda es una explicación, más lejos están de la experiencia inmediata las entidades a las que debe referirse, pero ello no significa que dichas entidades sean ficticias; bien al contrario, son parte de la propia estructura de la realidad.

Las explicaciones a menudo proporcionan predicciones, al menos, en principio. En efecto, si una cosa es, en principio, predecible, una explicación suficientemente completa de esa cosa deberá, en principio, hacer (entre otras cosas) predicciones completas sobre ella. Pero muchas cosas intrínsecamente impredecibles pueden también ser explicadas y entendidas. No podemos predecir, por ejemplo, qué números saldrán en una ruleta «limpia» (es decir, sin trampas), pero si comprendemos qué factores en el diseño y manejo de esa ruleta la hacen «limpia», podremos explicar por qué cualquier predicción resulta imposible. De nuevo, el mero conocimiento de que la ruleta está «limpia» no es lo mismo que comprender por qué lo está.

Así pues, será la comprensión, y no el mero conocimiento (o descripción, o predicción), el objeto de mi análisis. Dado que la comprensión llega gracias a las teorías explicativas, y dado que dichas teorías ofrecen a menudo amplias generalizaciones, la proliferación de datos registrados no hace necesariamente más difícil la comprensión de todo lo que es posible comprender. No obstante, la mayoría de las personas dirían —y es, en efecto, lo que me decían de pequeño— que no es sólo el número de datos registrados lo que ha ido incrementándose a un ritmo sobrecogedor, sino también la cantidad y complejidad de las teorías mediante las cuales comprendemos el mundo. En consecuencia —dirían—, aun suponiendo que hubo un tiempo en que una persona podía comprender todo lo que era comprensible entonces, ello resultaría ciertamente imposible hoy día y lo será más y más a medida que aumenten nuestros conocimientos. Y es que hay una impresión generalizada de que cada vez que se descubre una nueva explicación o técnica que es relevante en relación con un tema determinado, se ha de añadir una nueva

teoría a la lista, que debe aprender todo el que quiera comprender la materia en cuestión, y de que, cuando el número de teorías sobre una determinada materia aumenta demasiado, se desarrolla la especializacíón. La física, por ejemplo, se ha dividido en ciencias como la astrofísica, la termodinámica, la física de partículas, la teoría del campo cuántico y muchas más. Cada una de esas ciencias se basa en un entramado teórico tan rico, al menos, como el que tenía la física hace cien años, antes de ramificarse, y muchas de ellas ya se están fragmentando en nuevas subespecializaciones. Parece que cuanto más descubrimos, más irrevocablemente nos vemos metidos en la era del especialista y más lejos queda aquel tiempo en el que una persona podía aspirar, hipotéticamente, a comprender todo lo comprensible en su época.

Ya que nos enfrentamos a un menú de teorías producidas por la especie humana tan vasto y en tan rápido crecimiento, se nos puede perdonar que dudemos de que sea posible para un individuo no ya probar todos los platos en el espacio de una vida, sino ni siquiera tener noticia de todas las recetas existentes, al contrario de lo que, al parecer, resultaba factible antaño. Sin embargo, la explicación es una comida extraña: un bocado mayor no necesariamente cuesta más de tragar. Una teoría puede ser reemplazada por otra que explique más, sea más precisa y, además, resulte mas fácil de entender, en cuyo caso la vieja teoría resulta redundante, ganamos comprensión y no necesitamos aprender tanto como antes. Esto es lo que sucedió cuando la teoría de Copérnico de que la Tierra gira alrededor del Sol reemplazó al complejo sistema tolemaico, que situaba a la Tierra en el centro del universo. Una nueva teoría puede ser también una simplificación de otra existente, como ocurrió cuando la numeración arábiga (decimal) sustituyó a la romana. (En este caso, se trata de una teoría implícita. Un sistema de numeración que sustituye a otro anterior, por ser más sencillo, facilita ciertas operaciones y hace que surjan enunciados y maneras de pensar nuevos, lo cual conforma una teoría acerca de las relaciones entre los números.) Una nueva teoría puede surgir también de la unificación de otras anteriores y proporcionarnos una mayor comprensión que la que nos facilitaban éstas por separado, como sucedió cuando Michael Faraday y James Clerk Maxwell unieron las teorías de la electricidad y el magnetismo en una sola: la del electromagnetismo. De modo más indirecto, una mejora de las

explicaciones en cualquier campo tiende a mejorar las técnicas, los conceptos y el lenguaje con los que tratamos de comprender otros campos, con lo que nuestro conocimiento cuantitativo, además de alimentar, puede hacerse más comprensible desde un punto de vista estructural.

Sucede a menudo que aunque las teorías sean desplazadas por otras nuevas, no queden totalmente relegadas al olvido. Los números romanos aún se usan para determinados propósitos. Los engorrosos métodos que se utilizaban antiguamente para calcular (por ejemlo, que XIX veces XVII da CCCXXIII) no se aplican ya, pero siguen siendo conocidos y comprendidos por ciertas personas, como los historiadores de las matemáticas. ¿Significa esto que no podamos comprender «todo lo comprensible» sin conocer los números romanos y las complejas operaciones aritméticas que había que hacer para operar con ellos? Ciertamente, no. Un matemático moderno que jamás hubiese oído hablar de números romanos tendría, sin embargo, la capacidad para comprender perfectamente las matemáticas asociadas a ellos. Al estudiar los números romanos, este matemático no adquiriría una nueva comprensión, sino nuevos datos: datos históricos y datos acerca de las propiedades de ciertos símbolos definidos arbitrariamente, los cuales no aumentarían el conocimiento que ya tenía de los números propiamente dichos. Se encontraría en una situación similar a la de un zoólogo que aprendiese a traducir nombres de especies a un idioma extranjero, o a la de un astrofísico que aprendiese las diferentes maneras con las que las distintas culturas agrupan las estrellas formando constelaciones.

La situación sería distinta si el conocimiento de las operaciones aritméticas con números romanos pudiera ser necesario para la comprensión de la *historia*. Supongamos que una determinada teoría histórica —alguna explicación— dependiese de las técnicas específicas usadas por los antiguos romanos para multiplicar (una teoría semejante, por ejemplo, a la que postula que su técnica específica de fontanería, basada en la utilización de canalizaciones de plomo, que envenenaba el agua que bebían, pudo contribuir al declive del Imperio). En tal supuesto, deberíamos conocer en qué consistían esas técnicas para comprender la historia y, en consecuencia, todo lo comprensible.

En el caso que nos ocupa, sin embargo, ninguna explicación histórica depende de las técnicas de multiplicación, de modo que todo lo que nos interesa de ellas son meros enunciados de hechos. Todo lo comprensible puede ser comprendido sin estudiar esos hechos. Siempre podremos consultar las fuentes de referencia si, por ejemplo, tratamos de descifrar un texto antiguo que mencione dichas técnicas.

Al hacer hincapié en la distinción entre comprensión y «mero» conocimiento no pretendo menoscabar la importancia de la información no explicativa que poseemos. Por supuesto, es esencial para todo, desde la reproducción de un microorganismo (que la tiene codificada en sus moléculas de ADN) hasta el pensamiento humano más abstracto. Entonces ¿qué diferencia la comprensión del simple conocimiento? ¿Qué es una explicación, en oposición a un mero enunciado de hechos como puede ser una descripción o una predicción correctas? En la practica, reconocemos la diferencia con bastante facilidad. Sabemos cuándo no comprendemos algo, aunque podamos describirlo y predecirlo con certeza (el curso de una enfermedad conocida de origen desconocido, por ejemplo), y sabemos cuándo una explicación nos ayuda a comprenderlo mejor. Pero no resulta fácil dar una definición precisa de los términos «explicación» y «comprensión». Fundamentalmente ambos términos están más relacionados con los «porqué» que con los «qué», con el funcionamiento interno de las cosas, con cómo son éstas realmente y no con lo que aparentan ser, con lo que se desarrolla de acuerdo con una pauta más que con aquello que ocurre de modo aleatorio, con leyes de la naturaleza más que con estimaciones aproximadas. También se relacionan con la coherencia, la armonía y la simplicidad, en oposición a la arbitrariedad y la complejidad. Con todo, no son términos fáciles de definir. En cualquier caso, la comprensión es una función muy especial, una de las más elevadas de las que pueden llevar a cabo la mente y el cerebro humanos. Otros muchos sistemas físicos, como el cerebro de los animales, los ordenadores y diversas máquinas, pueden asimilar hechos y obrar en consecuencia, pero, hoy por hoy, no sabemos que haya nada capaz de comprender una explicación —o de buscarla, que tal vez sea lo más importante—, aparte de la mente humana.

El descubrimiento de nuevas explicaciones, y la comprensión de las explicaciones ya existentes, sólo están al alcance de la mente humana, la única que posee la facultad de pensar de manera creativa.

Cabría pensar que lo ocurrido con los números romanos es un proceso de «degradación» de una teoría explicativa a una mera descripción de hechos. Tales degradaciones ocurren de modo continuo a medida que crece nuestro conocimiento. Inicialmente, el sistema de numeración romano formaba parte del marco conceptual y teórico mediante el cual comprendían el mundo quienes lo usaban. Pero hoy el saber antaño representado por él sólo es una pequeña fracción del conocimiento, mucho más profundo, encarnado por las nuevas teorías matemáticas e, implícitamente, por sus modernos sistemas de numeración.

Todo lo cual ilustra otro atributo de la comprensión: es posible comprender algo sin darse cuenta de que se comprende, e incluso sin haber oído hablar específicameme de ello. Esto puede parecer paradójico, pero es evidente que una propiedad de las explicaciones profundas y generales es que cubren tanto situaciones familiares como desconocidas. Si fuésemos matemáticos modernos que nos encontrásemos por primera vez con los números romanos, quizás no nos daríamos cuenta al principio de que los comprendíamos. Primero tendríamos que saber a qué hechos se refieren, y luego considerar estos hechos a la luz de nuestro conocimiento actual de las matemáticas. Una vez realizado este proceso, tendríamos una perspectiva que nos permitiría comprender que lo único nuevo que representan para nosotros los números romanos es el mero hecho de su existencia. Es éste el sentido en que podemos decir que, a nivel explicativo, los números romanos son totalmente obsoletos.

De modo parecido, cuando digo que comprendo de qué manera la curvatura del tiempo y el espacio afecta la trayectoria de los planetas, incluso en otros sistemas solares de los que nunca he oído hablar, no pretendo ser capaz de dar automáticamente la explicación de todos los detalles del trayecto de cualquier órbita planetaria. Lo que quiero decir es que comprendo la teoría que contiene todas esas explicaciones, la cual, si se presenta la ocasión y dispongo de los datos relativos a un determinado planeta, me

permitirá calcular su órbita. Una vez calculada, podré mirar atrás y decir que no hay nada en el movimiento de dicho planeta, excepto meros hechos, que no esté explicado por la teoría general de la relatividad. Sólo comprendemos la estructura de la realidad si comprendemos las teorías que la explican, y puesto que éstas dan explicaciones que van más allá de lo que percibimos inmediatamente, podemos comprender cosas que están más allá de las que comprendemos de modo inmediato.

No estoy diciendo que cuando comprendemos una teoría comprendamos *necesariamente* todo lo que ésta puede explicar. A veces, cuando la teoría es muy profunda, llegar a la conclusión de que explica un determinado fenómeno puede constituir un descubrimiento significativo que requiera una explicación independiente. Por ejemplo, los quásares —fuentes de radiación extremadamente brillantes, situadas en el centro de algunas galaxias— constituyeron durante mucho tiempo uno de los misterios de la astrofísica. Se pensaba antaño que sería necesaria una nueva física para su explicación, mientras que ahora se considera que son explicados por la teoría general de la relatividad y otras teorías ya conocidas antes de que fueran descubiertos. Se supone que los quásares están formados por materia caliente en proceso de scaída dentro de agujeros negros (estrellas implosionadas cuyo campo gravitatorio es tan intenso que nada puede sustraerse a él). Llegar a esta conclusión, sin embargo, requirió años de investigación, tanto teórica como mediante la observación. Ahora que creemos haber avanzado algo en la comprensión de los quásares, no consideramos que ya la poseyéramos con anterioridad. La explicación de los quásares, aun mediante teorías ya existentes, nos ha aportado una comprensión genuinamente nueva. Resulta difícil definir qué es una explicación, y no lo es menos indicar cuándo una explicación subsidiaria se ha de considerar un componente individualizado del conocimiento y cuándo debe considerarse subsumida en la teoría profunda. Pero, aunque resulte difícil definirla, es relativamente sencillo reconocerla: en la práctica, nos damos cuenta de cuándo nos encontramos ante una nueva explicación. Una vez más, ello está relacionado con la creatividad. Describir el movimiento de un determinado planeta gracias a que comprendemos la explicación general que da la teoría de la gravitación universal no es más que una tarea mecánica, por compleja que pueda resultar, pero utilizar

las teorías que conocemos para explicar un fenómeno desconocido como los quásares requiere la intervención del pensamiento creativo. En resumen: para comprender todo lo comprensible hoy en astrofísica debemos conocer explícitamente la teoría de los quásares, mientras que no necesitamos saber la órbita específica de cada planeta.

Si bien el número de las teorías conocidas aumenta como una bola de nieve, al igual que la cantidad de hechos registrados, ello no implica necesariamente que la estructura total del conocimiento sea más difícil de comprender que antaño, pues aunque las teorías específicas son cada vez más numerosas y detalladas, también son «degradadas» de manera continua, a medida que el conocimiento que contienen se engloba en teorías profundas y generales. Por una parte, el número de estas teorías profundas y generales es cada vez menor, y, por otra, son cada vez más profundas y más generales. Que son «más profundas» significa que cada una de ellas explica más cosas —es decir, permite comprender más cosas— que las que la precedieron combinadas. Que son «más generales» significa que cada una de estas teorías dice más cosas, y acerca de un abanico más amplio de situaciones, que varias de las teorías independientes que existían con anterioridad.

Hace siglos, para construir una gran estructura, como una catedral o un puente, había que recurrir a un maestro de obras, que tenía un conocimiento práctico de lo que se requiere para dar a la estructura resistencia y estabilidad con el menor coste y esfuerzo posibles. No habría podido explicar sus conocimientos mediante el lenguaje de las matemáticas o la física, como sería el caso hoy día. En vez de ello, nuestro maestro de obras basaba sus conocimientos en una compleja trama de intuiciones, costumbres y reglas empíricas que había ido acumulando durante su aprendizaje y quizá mejorado gracias a la reflexión y a la experiencia. Con todo, esas intuiciones, costumbres y reglas empíricas eran, en realidad, teorías, explícitas o no, que contenían conocimientos reales de las especialidades que hoy día denominamos ingeniería y arquitectura. Y por esos conocimientos, tan modestos como se quiera comparados con los que poseemos hoy y con un abanico reducido de aplicaciones, se le habría contratado. Cuando admiramos obras centenarias, olvidamos a menudo que vemos

únicamente las sobrevivientes. La inmensa mayoría de las construcciones medievales y de épocas anteriores se hundieron hace tiempo, a menudo poco después de su construcción. Ello ocurría, sobre todo, con las estructuras innovadoras. En consecuencia, se daba por sentado que la innovación conllevaba el riesgo del desastre, y los constructores raramente se desviaban de los diseños y las técnicas validados por una larga tradición. Hoy día, en cambio, es muy raro que falle una estructura, aunque nunca haya sido ensayada previamente, por defectos de diseño. Cualquier obra que un constructor antiguo hubiese podido realizar, puede ser hecha hoy mejor y con mucho menor esfuerzo humano. Se pueden realizar también obras impensables en otros tiempos, por ejemplo, rascacielos y estaciones espaciales. Es posible utilizar materiales de los que un constructor antiguo jamás había oído hablar, como la fibra de vidrio y el hormigón armado, y que si, por un milagro, se viese trasladado a nuestra época, no sabría usar por falta de una adecuada comprensión de su funcionamiento.

El progreso hasta llegar al estado actual de conocimientos no se consiguió mediante la acumulación de más teorías como las que utilizaba el maestro de obras. Nuestro conocimiento, tanto explícito como implícito, no sólo es mucho mayor, sino también distinto estructuralmente. Como he dicho, las teorías modernas son menos en número, pero más generalesy más profundas. Para cada caso al que se enfrentaba —por ejemplo, decidir el espesor de un muro de carga— nuestro constructor disponía de un repertorio de intuiciones específicas y reglas empíricas que, sin embargo, podían aportar soluciones completamente erróneas si se aplicaban a situaciones nuevas. En la actualidad, deducimos nuestras respuestas mediante una teoría lo suficientemente general para ser aplicada a paredes de cualquier material en cualquier circunstancia: en la Luna, bajo el agua o donde sea. La razón de que pueda aplicarse de modo tan general es que se basa en un profundo conocimiento de la manera en que trabajan materiales y estructuras. Para hallar el adecuado espesor de un muro de carga que se ha de construir con un material poco familiar, emplearemos la teoría utilizada para calcular todos los muros de carga, pero basaremos nuestros cálculos en el hecho de que se trata de un material diferente, de modo que usaremos valores diferentes para los distintos parámetros. Hay que tomar en consideración hechos tales como la resistencia tensorial y la

elasticidad del material, pero, por lo demás, no necesitaremos comprensión adicional alguna.

Ésta es la razón por la que, a pesar de tener una comprensión mucho mayor que cualquier maestro de obras de la antigüedad, un arquitecto moderno no necesita una preparación más larga ni más difícil. Es posible que las teorías de un plan de estudios actual resulten más difíciles de comprender que las reglas empíricas del antiguo constructor, pero dado que el número de teorías actuales es relativamente reducido y su poder explicativo les otorga cualidades adicionales, como la belleza, la lógica interna y las conexiones con otras materias, resultan más fáciles de aprender. Ahora sabemos que algunas de las antiguas reglas empíricas eran erróneas, otras eran ciertas y otras eran aproximaciones bastante razonables a la realidad, y, además, sabemos el porqué. Unas pocas siguen vigentes, pero ninguna de ellas es ya la base de la comprensión de lo que mantiene en pie a las estructuras para ningún arquitecto.

No niego, por supuesto, que se dé la especialización en muchas áreas en las que el conocimiento está creciendo, incluyendo la arquitectura. Pero no es un proceso unidireccional, ya que las especializaciones también desaparecen a menudo: los arados y las ruedas ya no son diseñados y manufacturados por artesanos, ni las cartas escritas por amanuenses. Y no resulta menos evidente que la tendencia profundizadora y unificadora que he descrito no es la única que se da hoy día, puesto que simultáneamente hay una continua *ampliación*, es decir, que las nuevas ideas hacen a menudo algo más que reemplazar, simplificar o unificar las existentes: amplían también la comprensión humana hacia áreas mal comprendidas con anterioridad, o cuya existencia ni siquiera se sospechaba. Pueden así ofrecer nuevas oportunidades, nuevas especializaciones y nuevas materias, así como plantear nuevos problemas; y cuando esto sucede, nos encontramos —al menos temporalmente— con que tenemos que ampliar nuestros conocimientos para mejorar nuestra comprensión.

La ciencia de la medicina es quizá la citada más a menudo como ejemplo de creciente especialización, consecuencia, al parecer inevitable, de su incesante crecimiento a medida que se

descubren nuevos procesos de curación y mejores tratamientos para más enfermedades. Pero incluso en la medicina se manifiesta, cada vez más, la tendencia unificadora opuesta. Sin duda, muchas de las funciones del cuerpo no se conocen todavía bien, al igual que los mecanismos de numerosas enfermedades. En consecuencia, algunas áreas del conocimiento médico siguen basándose en la experiencia y el ojo clínico de facultativos expertos en determinadas enfermedades y tratamientos, que transmiten esa experiencia y ese ojo clínico a las siguientes generaciones de médicos. Dicho de otro modo, gran parte de la medicina sigue en la era del conocimiento empírico, y cuando surgen nuevos conocimientos empíricos aumentan los incentivos para la especialización. Pero como la investigación médica y bioquímica ofrece sin cesar explicaciones más profundas y completas sobre los procesos de enfermedad y curación, la comprensión también va en aumento. Cada vez hay más conceptos específicos que son reemplazados por otros generales a medida que se descubren mecanismos moleculares subyacentes comunes a enfermedades semejantes en distintas partes del cuerpo. En cuanto una enfermedad puede ser situada dentro de un marco general, el papel de los especialistas disminuye. Y, paralelamente, los médicos que se encuentran ante una enfermedad desconocida o una complicación inesperada tienen cada vez más posibilidades de relacionarlas con teorías explicativas. Pueden consultar la bibliografía acerca de los hechos a los que se enfrentan y luego aplicar una teoría general al caso específico que les ocupa, a fin de desarrollar un tratamiento específico con la esperanza de que resulte efectivo, aunque no se haya utilizado con anterioridad.

De este modo, la cuestión de si se está volviendo más difícil o más fácil comprender todo lo comprensible dependerá del equilibrio final entre los dos efectos opuestos del crecimiento del conocimiento: la *ampliación* de nuestras teorías y su creciente *profundidad*. Que se amplíen aumenta la dificultad, mientras que el hecho de que se profundicen la disminuye. Una de las tesis de este libro es que despacio, pero con firmeza, la profundidad va ganando terreno. Dicho de otro modo, en la práctica, la afirmación que me resistía a aceptar cuando era niño se confirma como falsa y la opuesta a ella como cierta.

No nos alejamos del estado en que una persona podría comprender todo lo comprensible, sino que nos acercamos a él.

No es que estemos a punto de comprenderlo *todo*. Se trata de una cuestión muy diferente. No creo que estemos, ni que lleguemos a estar nunca, a punto de comprender *todo lo que existe*. Estoy hablando de posibilidad de comprender *todo lo comprensible*. Ello depende más de la estructura de nuestro conocimiento que de su contenido; pero, sin duda, la estructura de nuestro conocimiento —el hecho de que sea expresable o no por medio de teorías que encajen formando un todo comprensible— depende de cuál sea la naturaleza de la estructura total de la realidad. Si el conocimiento ha de seguir creciendo de un modo que, aparentemente, no tiene límites, y, a pesar de ello, nos encaminamos hacia un estado en el que sea posible para los seres humanos comprender todo lo comprensible, la profundidad de nuestras teorías debe crecer con la suficiente rapidez para que resulte factible. Y ello sólo será así si la estructura de la realidad está en sí misma altamente unificada, de modo que pueda ser comprendida cada vez más a medida que aumente nuestro conocimiento. Si es así, nuestras teorías llegarán a ser tan generales, y profundas, y a estar tan integradas entre sí, que se convertirán, de hecho, en una sola teoría de una estructura unificada de la realidad. Esta teoría seguirá sin explicar *todos* los aspectos de la realidad, aspiración inalcanzable, pero englobará *todas* las explicaciones conocidas y será de aplicación a la estructura total de la realidad hasta el punto en que ésta resulte comprensible. Mientras que las teorías precedentes se referían a temas determinados, ésta será una teoría de todos los temas: una *teoría total de la realidad*.

No será, por supuesto, la teoría definitiva, sino la primera de una serie de teorías cada vez más perfectas. En ciencia, damos por sentado que aun nuestras mejores teorías actuales están condenadas a ser imperfectas y problemáticas en algunos aspectos, y esperamos que sean reemplazadas en su momento por otras más profundas y exactas. Este progreso no se detiene con el descubrimiento de una teoría universal. Newton, por ejemplo, nos proporcionó la primera teoría universal de la gravitación, que conllevó la unificación, entre otras ramas de la física, de la mecánica celeste y la terrestre, pero sus teorías se vieron

reemplazadas por la teoría general de la relatividad de Einstein, que incorpora además la geometría (considerada con anterioridad una rama de las matemáticas) a la física, con lo que proporciona explicaciones más profundas y exactas. La primera teoría plenamente universal —la que denomino teoría total de la realidad— no será, al igual que las que la precedieron y las que la seguirán, ni perfectamente cierta ni infinitamente profunda, de modo que, en su momento, también será reemplazada. Pero no lo será mediante unificaciones con teorías sobre otras materias, puesto que será ya una teoría de todas las materias. En el pasado, algunos grandes avances en la comprensión nacieron de grandes unificaciones. Otros fueron consecuencia de cambios estructurales en el modo de comprender una determinada materia, como cuando se dejó de considerar que la Tierra era el centro del universo. Tras la teoría total de la realidad no habrá más grandes unificaciones. Todos los grandes descubrimientos subsiguientes conllevarán cambios en nuestra forma de comprender el mundo como un todo: serán cambios en nuestra concepción del mundo. La consecución de una teoría total de la realidad constituirá la última gran unificación y, al mismo tiempo, el primer salto hacia una nueva concepción del mundo. Creo que esta unificación y este salto se están operando actualmente, y la concepción del mundo asociada a ellos constituye el tema de este libro.

Debo aclarar, antes de seguir, que no me refiero a la mera «teoría total» que algunos físicos de partículas aspiran a descubrir en breve. Su «teoría total» sería, en realidad, una teoría unificada de todas las fuerzas básicas conocidas en física, como la gravedad, el electromagnetismo y las fuerzas nucleares. Describiría también todos los tipos de partículas subatómicas existentes, sus masas, sus espines, sus cargas eléctricas y demás propiedades, así como el modo en que interactúan. Dada una descripción suficientemente precisa del estado inicial de un sistema físico aislado, podría, en principio, predecir su comportamiento. Cuando el comportamiento exacto de un sistema resultase impredecible, describiría todos los comportamientos posibles y sus probabilidades. En la práctica, ocurre a menudo que los estados iniciales de sistemas interesantes no pueden ser determinados con exactitud y, por otra parte, el cálculo de las predicciones resultaría demasiado complicado, excepto en los casos más simples.

No obstante, semejante teoría unificada de partículas y fuerzas, junto con una especificación del estado inicial del universo en el momento del Big Bang (la violenta explosión que lo originó), contendría, en principio, toda la información necesaria para predecir todo lo predecible (figura 1.1).

Pero predicción no equivale a explicación. La tan deseada «teoría total», incluso combinada con una teoría del estado inicial, no proporcionaría, en el mejor de los casos, más que una pequeña faceta de una verdadera teoría total de la realidad. Podría predecirlo todo (en principio), pero no es de esperar que pudiese explicar mucho más que las teorías existentes, excepto en el caso de unos pocos fenómenos dominados por las particularidades de las interacciones subatómicas, tales como las colisiones en el interior de los aceleradores de partículas y la exótica historia de las transmutaciones de partículas en el Big Bang. ¿Qué motiva la utilización del pomposo término «teoría total» para designar un fragmento tan limitado del conocimiento, por muy fascinante que sea? Se trata, en mi opinión, una vez más, de una visión errónea de la naturaleza de la ciencia, censurada por muchos críticos de la ciencia y alabada (lamentablemente) por otros tantos científicos, que sostiene que la ciencia es, en esencia, reduccionista, es decir, que, supuestamente, analiza las cosas de manera reductiva, mediante el estudio de sus componentes. La resistencia de un muro a la penetración o al derrumbamiento, por ejemplo, es explicada contemplándolo como una vasta agregación de moléculas interactuantes. Las propiedades de estas moléculas se explican, a su vez, en términos de sus átomos constituyentes y de las interacciones que los vinculan, y así sucesivamente, hasta llegar a las más ínfimas partículas y las fuerzas más básicas. Los reduccionistas creen que toda explicación científica, e incluso cualquier explicación, en general, que sea lo suficientemente profunda, debe estar estructurada de este modo.

La concepción reduccionista conduce de modo natural a una clasificación jerárquica de materias y teorías, de acuerdo con su proximidad a las teorías predictivas de más «bajo nivel» conocidas. En esta jerarquía, la lógica y las matemáticas constituyen la roca inamovible sobre la que se construye el edificio de la ciencia. La piedra angular sería una «teoría total» reductora, una teoría universal de partículas, fuerzas, espacio y tiempo, junto con una teoría sobre el estado inicial del universo. El resto de la física constituiría las primeras plantas del edificio, mientras que la astrofísica y la química se situarían en un nivel superior, la geología aún más arriba, y así sucesivamente. El edificio se ramifica en múltiples torres de materias de un nivel cada vez más alto, como la bioquímica, la biología y la genética. Encaramadas en la parte más alta, en la estratosfera, por así decirlo, encontraríamos la teoría de la evolución, la economía, la psicología y la informática, que en esta imagen resultan casi inconcebiblemente secundarias.

Por el momento, disponemos únicamente de algunas aproximaciones a una «teoría total» reductora, capaces de predecir con bastante exactitud leyes del movimiento de partículas subatómicas individuales. Partiendo de estas leyes, los ordenadores actuales pueden calcular con cierto detalle el movimiento de cualquier grupo de unas cuantas partículas interactuantes, conocido su estado inicial. Sin embargo, hasta el más pequeño fragmento de materia perceptible a simple vista contiene trillones de átomos, cada uno de ellos compuesto por muchas partículas subatómicas que interactúan de modo constante con el mundo exterior, de modo que resulta prácticamente irrealizable la predicción de su comportamiento partícula por partícula. Complementando las leyes exactas del movimiento con algunos esquemas de aproximación, podremos predecir ciertos aspectos del comportamiento general de objetos relativamente grandes, como, por ejemplo, la temperatura a la que un determinado compuesto químico se fundirá o hervirá. Gran parte de la química básica ha sido reducida de este modo a física. Pero para las ciencias de nivel superior el planteamiento reduccionista queda limitado a una cuestión de principio. Nadie espera realmente deducir muchos principios de biología, psicología o política partiendo de los de la física. La razón por la cual las materias de nivel superior pueden ser estudiadas es que, en

circunstancias especiales, el abrumadoramente complejo comportamiento de un vasto número de partículas se reduce por sí mismo a una dimensión de simplicidad y comprensibilidad accesibles. Esto recibe el nombre de *emergencia*: la simplicidad de alto nivel «emerge» de la complejidad de bajo nivel. Los fenómenos de alto nivel sobre los que disponemos de hechos comprensibles, no deducibles simplemente de teorías de nivel inferior, reciben el nombre de *fenómenos emergentes*. Un muro, por ejemplo, puede ser resistente porque sus constructores temiesen que sus enemigos tratasen de abrir una brecha en él. Esta es una explicación de alto nivel para la resistencia del muro, no deducible de la explicación de bajo nivel que ofrecí antes (aunque no incompatible con ella). «Constructores», «enemigos», «temor» y «tratar» son todos ellos fenómenos emergentes. El propósito de las ciencias de alto nivel es capacitarnos para la comprensión de los fenómenos emergentes, los más importantes de los cuales son, como veremos, *vida, pensamiento y calculabilidad*.

Por cierto, lo contrario del reduccionismo, es decir, el *holismo*—la idea de que las únicas explicaciones legítimas son las desarrolladas en términos de sistemas de nivel superior—, constituye un error aún mayor que aquél. ¿Qué esperan los holistas que hagamos? ¿Dejar de investigar las causas moleculares de las enfermedades? ¿Negar que los seres humanos estamos formados por partículas subatómicas? Cuando existe una explicación reduccionista, es tan deseable como cualquier otra. Cuando ciencias enteras son reducibles a ciencias de nivel inferior, nos incumbe a los científicos averiguar esas reducciones tanto como descubrir cualquier otro conocimiento.

Los reduccionistas piensan que la tarea de la ciencia consiste en analizar las cosas por sus componentes. Los instrumentalistas creen que consiste en predecir sucesos. Para los miembros de ambas corrientes la existencia de ciencias de alto nivel es meramente una cuestión de conveniencia. La complejidad nos impide la utilización de la física fundamental para realizar predicciones de alto nivel, de modo que, en vez de ello, aventuramos lo que serían esas predicciones si pudiésemos realizarlas —la emergencia nos brinda la posibilidad de hacerlo con éxito—, y ésta es, supuestamente, la tarea de las ciencias. Así

pues, para reduccionistas e instrumentalistas, que desprecian la verdadera estructura y el verdadero propósito del conocimiento científico, la base de la jerarquía predictiva de la física es, por definición, la «teoría total». Pero para todos los demás científicos el conocimiento científico consiste en explicaciones, y la estructura de la explicación científica no refleja la jerarquía reduccionista. Hay explicaciones en cada nivel de la jerarquía, muchas de las cuales son autónomas, referidas únicamente a conceptos de su nivel específico (por ejemplo, «el oso se comió la miel porque estaba hambriento»). Muchas de ellas comprenden deducciones en el sentido opuesto al de la explicación reductiva, es decir, no explican las cosas mediante el análisis de sus partes, sino contemplándolas como partes de un todo mayor y más complejo, para el cual disponemos, no obstante, de teorías explicativas. Consideremos, por ejemplo, una determinada molécula de cobre de la estatua de Sir Winston Churchill en la plaza del Parlamento de Londres. Intentaré explicar por qué está ahí esa molécula de cobre: porque Churchill fue diputado a la cercana Cámara de los Comunes y primer ministro, porque sus ideas y liderazgo contribuyeron a la victoria de los Aliados en la Segunda Guerra Mundial, porque es costumbre honrar a tales personas erigiéndoles estatuas, porque éstas suelen realizarse en bronce y porque esta aleación contiene cobre. Explicamos así una observación física de bajo nivel —la presencia de una molécula de cobre en un determinado lugar—, mediante teorías de nivel extremadamente alto sobre fenómenos emergentes tales como ideas, liderazgo, guerra y tradición.

No hay razón alguna para que exista, ni siquiera en principio, otra *explicación* de nivel inferior para la presencia de esa molécula de cobre que la que he ofrecido. Es presumible que una «teoría total» reactiva ofreciera, en principio, una *predicción* de bajo nivel de la probabilidad de que semejante estatua existiese, dada la condición (por ejemplo) del sistema solar en una determinada fecha anterior. Describiría también, en principio, cómo había llegado probablemente la estatua a su emplazamiento. Pero tales descripciones y predicciones (además de ser inútiles, por supuesto) no nos explicarían nada. Nos describirían tan sólo la trayectoria que siguió cada átomo de cobre desde la mina hasta el estudio del escultor pasando por la fundición, y así sucesivamente. Dejarían constancia, también, de cómo estas trayectorias se vieron influidas por las fuerzas ejercidas por los átomos circundantes, por

ejemplo, los de los cuerpos de los mineros y del escultor, y así predecirían la existencia de la escultura y su forma. En realidad, semejante predicción debería referirse a la totalidad de los átomos del planeta involucrados en la compleja acción que denominamos Segunda Guerra Mundial, entre otras cosas. Pero, incluso si dispusiéramos de la sobrehumana capacidad necesaria para seguir tan largas predicciones sobre la situación del átomo en cuestión, no estaríamos en condiciones de afirmar que entendemos por qué está donde se halla. Sabríamos, meramente, que su llegada y su modo de llegar eran inevitables (o de esperar, o lo que se quiera), dadas las configuraciones iniciales de todos los átomos y las leyes de la física. Para saber el porqué, deberíamos necesariamente dar un paso más. Tendríamos que investigar qué había en aquella configuración de átomos y en aquellas trayectorias que hizo que tendieran a depositar aquella determinada molécula de cobre en aquella precisa posición. Realizar esta investigación constituiría una labor creativa, como es siempre el descubrimiento de nuevas explicaciones. Deberíamos descubrir que determinadas configuraciones atómicas favorecen que surjan fenómenos emergentes, tales como el liderazgo y la guerra, relacionados entre sí por teorías explicativas de alto nivel. Sólo cuando conociésemos estas teorías podríamos comprender plenamente por qué la molécula de cobre está donde se halla.

En la concepción del mundo reduccionista, las leyes que gobiernan las interacciones de las partículas subatómicas son de la mayor importancia, puesto que constituyen la base de la jerarquía de todo el conocimiento. Sin embargo, tanto en la verdadera estructura del conocimiento científico como en la de nuestro conocimiento en general, dichas leyes tienen un papel mucho más humilde.

¿Cuál es ese papel? Para mí, ninguna de las teorías postuladas hasta ahora como candidatas a «teoría total» contiene demasiadas novedades en lo que se refiere a la explicación. Quizás el enfoque más novedoso, desde el punto de vista explicativo, sea el de *la teoría de la supercuerda*, en la que objetos alargados, o «cuerdas», en vez de partículas puntuales, constituyen los elementos fundamentales de la materia. Pero lo cierto es que ningún enfoque actual parece capaz de ofrecer un modo de explicación enteramente nuevo; nuevo en el sentido de la

explicación de Einstein de las fuerzas gravitatorias en términos de tiempo y espacio curvos. De hecho, se espera que la futura «teoría total» herede virtualmente toda su estructura explicativa—sus conceptos físicos, su lenguaje, su formalismo matemático y la forma de sus explicaciones— de las actuales teorías del electromagnetismo, las fuerzas nucleares y la gravedad. Podemos, pues, buscar en esta estructura subyacente, que conocemos gracias a las teorías ya existentes, la aportación de la física fundamental a nuestra comprensión general.

Existen en física dos teorías considerablemente más profundas que las demás. La primera es la teoría general de la relatividad, que, como he dicho, es la mejor acerca del espacio, el tiempo y la gravedad. La segunda, la *teoría cuántica*, es aún más profunda. Entre las dos, estas teorías proporcionan (cosa que no hace ninguna de las teorías de partículas subatómicas existentes o propuestas) el detallado entramado explicativo y formal dentro del que se expresan todas las restantes teorías de la física moderna, y contienen principios físicos de tan amplio alcance, que todas las demás teorías se ajustan a ellos. Desde hace décadas, la unificación de la teoría general de la relatividad y la teoría cuántica —que daría una *teoría cuántica de la gravedad*— es un reto para los físicos teóricos, y debería formar parte de cualquier «teoría total», así como de la teoría total de la realidad. Como veremos en el próximo capítulo, la teoría cuántica, al igual que la de la relatividad, proporciona un modo de explicación nuevo y revolucionario de la realidad física. La razón de que la teoría cuántica sea la más profunda de las dos se encuentra más bien fuera de la física que dentro de ella, pues sus ramificaciones son muy amplias y se extienden más allá de la física e incluso más allá de la ciencia, tal como la concebimos normalmente. La teoría cuántica constituye una de las *cuatro vías principales* que, en mi opinión, conforman nuestro conocimiento actual de la estructura de la realidad.

Antes de exponer cuáles son las tres vías restantes, debo mencionar otra de las maneras en que el reduccionismo interpreta erróneamente la estructura del conocimiento científico. No sólo asume que la explicación siempre consiste en analizar un sistema fragmentándolo en sistemas más pequeños y sencillos, sino también que toda explicación debe basarse en hechos posteriores

y realizarse mediante hechos anteriores; en otras palabras, que el único modo de explicar algo es precisar sus *causas*. Ello implica que cuanto más anteriores sean los hechos en cuyos términos intentamos explicar algo, mejor será la explicación, de modo que, en última instancia, la explicación más perfecta se encuentra en el estado inicial del universo.

Una «teoría total» que excluya la especificación del estado inicial del universo no es una descripción completa de la realidad física porque proporciona únicamente leyes del movimiento, y éstas, por sí mismas, sólo realizan predicciones condicionales. Es decir, nunca afirman categóricamente lo que sucede, sino lo que sucederá en un momento determinado dado lo que sucedía en otro momento determinado. Sólo si se proporciona una especificación completa del estado inicial puede, en principio, deducirse una descripción completa de la realidad física. Las teorías cosmológicas actuales no proporcionan una especificación completa del estado inicial ni tan sólo en principio, pero afirman que el universo era en sus inicios muy reducido, estaba caliente y tenía una estructura muy uniforme. Sabemos también que no podía ser perfectamente uniforme, ya que ello resultaría incompatible, según la teoría, con la distribución de galaxias que observamos hoy en el cielo. Las diferencias iniciales en densidad debieron de incrementarse notablemente a causa de la densificación gravitatoria (es decir, las regiones relativamente densas atrajeron a más materia que las otras y se densificaron aún más), de modo que pudieron haber sido muy pequeñas al principio. Sin embargo, y por pequeñas que fuesen, esas diferencias tienen una importancia fundamental en cualquier descripción reduccionista de la realidad, pues casi todo lo que podemos observar a nuestro alrededor, desde la distribución de galaxias y estrellas en el cielo hasta la aparición de estatuas de bronce en el planeta Tierra, es, desde el punto de vista de la física fundamental, consecuencia de dichas variaciones.

Si nuestra descripción reduccionista ha de explicar algo más que las características más generales del universo observado, necesitamos una teoría que especifique esas cruciales desviaciones iniciales de la uniformidad.

Permítaseme que reformule esta exigencia sin el condicionamiento reduccionista. Las leyes del movimiento, para cualquier sistema, hacen únicamente predicciones condicionales y son, por tanto, compatibles con múltiples posibles historias de ese sistema. (Este planteamiento es independiente de las limitaciones impuestas a la predecibilidad por la teoría cuántica, que analizaré en el próximo capítulo.)

Por ejemplo, las leyes del movimiento que rigen a un proyectil disparado desde un cañón son compatibles con múltiples trayectorias, una para cada posible dirección y elevación en que se encontrase el arma en el momento del disparo (figura 1.2). Matemáticamente, las leyes del movimiento pueden expresarse mediante una serie de ecuaciones denominadas *ecuaciones del movimiento*, que arrojan múltiples soluciones distintas, una para cada posible trayectoria.

Para especificar qué solución describe la trayectoria real, debemos aportar *datos suplementarios* sobre lo que realmente sucede. Un modo de hacerlo es especificar el estado inicial, en este caso, la dirección en la que apunta el cañón. Pero existen otras maneras. Podemos, por ejemplo, especificar el estado final, es decir, la posición y dirección del movimiento del proyectil en el momento de hacer blanco. O podemos especificar la posición del punto más alto de la trayectoria. No importa qué datos complementarios aportemos, mientras se ajusten a alguna de las posibles soluciones de las ecuaciones del movimiento. La combinación de cualquiera de esos datos con las leyes del movimiento desemboca en una teoría que describe todo lo que le sucede al proyectil desde el lanzamiento hasta el impacto.

De modo semejante, las leyes del movimiento para la realidad física, considerada en su totalidad, tendrían múltiples soluciones, cada una de las cuales correspondería a una historia distinta. Para completar la descripción, deberíamos especificar cuál de las historias posibles es la que realmente ocurrió y aportar suficientes datos complementarios para que sea posible una de las múltiples soluciones que ofrecen las ecuaciones del movimiento. Una manera de aportar estos datos, al menos cuando se trata de modelos cosmológicos sencillos, es especificar el estado inicial del universo, pero podríamos especificar igualmente su estado final, o el que tiene en un determinado momento, o incluso aportar información sobre el estado inicial, el final y los intermedios. En general, la combinación de suficientes datos de cualquier tipo con las leyes del movimiento nos daría una descripción completa, en principio, de la realidad física.

Por lo que se refiere al proyectil de cañón, una vez especificado, por ejemplo, el estado final, el cálculo del estado inicial es directo, y viceversa, de modo que no existe diferencia práctica entre modos distintos de especificar datos complementarios. Sin embargo, cuando se trata del universo, la mayoría de estos cálculos son impracticables. Como he dicho, inferimos la existencia de una densificación en las condiciones iniciales por la observación de la densificación actual, pero eso es algo excepcional: la mayor parte de nuestro conocimiento sobre datos complementarios —de lo que realmente sucede— se basa en teorías de alto nivel sobre fenómenos emergentes, y, en consecuencia, por definición, no puede expresarse de manera práctica en forma de afirmaciones sobre el estado inicial. Por ejemplo, en la mayoría de las soluciones de las ecuaciones del movimiento el estado inicial del universo carece de las condiciones adecuadas para el desarrollo de la vida. Por lo tanto, nuestro conocimiento de que la vida se ha desarrollado constituye un elemento fundamental en los datos complementarios. Quizás nunca sepamos qué implica específicamente esta restricción respecto de la detallada estructura del Big Bang, pero podemos sacar de ella conclusiones directamente. Por ejemplo, la primera estimación precisa sobre la edad de la Tierra fue hecha de acuerdo con la teoría biológica de la evolución, que contradecía la mejor física del momento.

Sólo un prejuicio reduccionista podría hacernos creer que esa manera de razonar es menos válida que las otras o que, en general, es más «fundamental» una teoría sobre el estado inicial que una sobre las características emergentes de la realidad.

Incluso en el dominio de la física fundamental, la idea de que las teorías del estado inicial contienen nuestro más profundo conocimiento resulta completamente errónea. Una razón es que, lógicamente, dicha idea excluye la posibilidad de explicar el propio estado inicial —por qué era lo que era—, cuando lo cierto es que disponemos de descripciones sobre múltiples aspectos de dicho estado. Y, de modo más general, es imposible que ninguna teoría del *tiempo* lo explique en términos de nada «anterior», mientras que disponemos, gracias a la teoría de la relatividad general y, sobre todo, a la teoría cuántica, de profundas explicaciones de la naturaleza del tiempo (véase capítulo 11).

Así pues, muchas de nuestras descripciones, predicciones y explicaciones de la realidad no tienen ningún parecido con la imagen «estado inicial más leyes del movimiento» a la que nos conduce el reduccionismo. No existe razón alguna para considerar a las teorías de alto nivel como «ciudadanas de segunda». Las teorías de la física subatómica, e incluso la teoría cuántica y la de la relatividad, no son superiores, ni mucho menos, a las teorías sobre las propiedades emergentes. No es posible que ninguna de esas áreas del conocimiento subsuma a las otras. Cada una de ellas tiene implicaciones lógicas para las demás, pero no todas estas implicaciones pueden ser establecidas, puesto que constituyen propiedades emergentes de los dominios de las otras teorías. En realidad, incluso los términos «alto nivel» y «bajo nivel» resultan engañosos. Las leyes de la biología, por ejemplo, son de alto nivel, consecuencias emergentes de las leyes de la física, pero, lógicamente, algunas de las leyes de la física son entonces consecuencias «emergentes» de las leyes de la biología. Podría incluso ser que, entre ellas, las leyes que rigen los fenómenos biológicos y otros fenómenos emergentes, determinasen por completo las leyes fundamentales de la física. Pero, en cualquier caso, cuando dos teorías están lógicamente relacionadas, la lógica no dicta cuál de ellas deberíamos considerar determinante, de modo total o parcial, de la otra. Ello dependerá

de las relaciones explicativas entre ellas. Las teorías verdaderamente superiores no son las que se refieren a un determinado nivel de tamaño o complejidad, ni las situadas en un determinado nivel de la jerarquía predictiva, sino las que contienen las explicaciones más profundas. La estructura de la realidad no consiste únicamente en ingredientes reduccionistas como espacio, tiempo y partículas subatómicas, sino también en vida, pensamiento, calculabilidad y las demás materias a las que se refieren dichas explicaciones. Lo que convierte a una teoría en más fundamental y menos secundaria no es su proximidad a la supuesta base predictiva de la física, sino a las más profundas teorías explicativas.

La teoría cuántica, como ya he dicho, es una de esas teorías, y las otras tres vías explicativas por medio de las cuales intentamos comprender la estructura de la realidad son de «alto nivel» desde el punto de vista de la física cuántica. Son la *teoría de la evolución* (básicamente, de la evolución de los organismos vivos), la *epistemología* (la teoría del conocimiento) y la *teoría de la calculabilidad* (que trata de los ordenadores y lo que, en principio, pueden calcular o no). Como demostraré, se han hallado tantas y tan diversas conexiones entre los principios básicos de estas cuatro materias, aparentemente independientes, que resulta imposible conseguir una auténtica comprensión de cualquiera de ellas sin comprender al mismo tiempo las otras tres. Las cuatro juntas constituyen una estructura explicativa coherente de amplísimo alcance y que abarca una porción muy considerable de nuestro conocimiento del mundo, por lo que, en mi opinión, puede ya ser con justicia denominada la primera verdadera teoría de total de la realidad. Hemos llegado así a un momento significativo en la historia de las ideas: el momento en que el alcance de nuestra comprensión empieza a ser verdaderamente universal. Hasta ahora, toda nuestra comprensión giraba sobre algún aspecto de la realidad que se destacaba del conjunto de su atipicidad. En el futuro, lo hará sobre una visión unificada de la realidad: todas las explicaciones serán comprendidas desde la óptica de la universalidad, y cada nueva idea tenderá automáticamente a iluminar no sólo un determinado aspecto, sino, en grados diversos, *todos* los aspectos.

El aumento de la comprensión que con el tiempo puede proporcionar esta última gran unificación es posible que supere con mucho el proporcionado por cualquiera de las anteriores.

Como veremos, no es sólo la física la que es explicada y unificada por esa teoría, y no tan sólo la ciencia, sino también, potencialmente, los límites más lejanos de la filosofía, la lógica, las matemáticas, la ética, la política y la estética; quizá todo lo que comprendemos en la actualidad y, probablemente, mucho de lo que aún no comprendemos.

¿Qué conclusión pues, podría ofrecer a aquel niño que fui antaño, que rechazaba la posibilidad de que el crecimiento del conocimiento hiciera que el mundo fuera menos comprensible? Estaría de acuerdo con él, si bien ahora creo que lo importante no es que lo que resulte comprensible para nuestra especie pueda ser comprendido por *cualquiera* de sus miembros, sino que la estructura de la realidad esté verdaderamente unificada y pueda comprenderse. Todo induce a creer que es así. Cuando era niño, simplemente, lo sabía; ahora puedo explicarlo.

TERMINOLOGÍA

Epistemología. Estudio de la naturaleza del conocimiento y los procesos que lo crean.

Explicación. Someramente, enunciado que describe la naturaleza de las cosas y su razón de ser.

Instrumentalismo. Corriente del pensamiento científico que postula que las teorías científicas deben limitarse a predecir correctamente el resultado de los experimentos.

Positivismo. Forma extrema de instrumentalismo, que sostiene que todo enunciado que no describa o prediga observaciones es irrelevante. (Este punto de vista también resulta irrelevante, de acuerdo con sus propios criterios.)

Reduccionismo. Corriente del pensamiento científico que postula que la ciencia debe explicar los hechos analizando sus componentes de nivel inferior.

Holismo. Es la corriente opuesta al reduccionismo, y postula que la ciencia debe explicar los hechos analizando sus componentes de nivel superior.

Fenómeno emergente. Es aquel (como la vida, el pensamiento o la calculabilidad) acerca del cual existen hechos comprensibles o explicaciones que no son simples deducciones lógicas de teorías de nivel inferior, sino que pueden comprenderse o predecirse mediante teorías de nivel superior directamente referidas a cada caso concreto.

SUMARIO

El conocimiento científico, como todo conocimiento humano, se basa principalmente en explicaciones. Los meros hechos pueden ser consultados en las fuentes, mientras que las predicciones son importantes únicamente a fin de efectuar pruebas experimentales cruciales para discriminar entre las teorías científicas en competición que han superado ya la prueba de ser buenas explicaciones. A medida que nuevas teorías arrinconan a las que las precedieron, nuestro conocimiento se amplía (con la incorporación de nuevas materias) y se profundiza (al explicar más las teorías fundamentales y hacerse más generales). La profundidad va en aumento, de modo que no sólo no nos alejamos del estado en que una persona podría comprender todo lo comprensible, sino que nos encaminamos hacia él. Las teorías más profundas se integran de tal manera unas con otras, que sólo pueden ser comprendidas conjuntamente, como una única teoría de la estructura unificada de la realidad. Esta teoría total de la realidad tiene un alcance infinitamente mayor que la «teoría total» que buscan los físicos de partículas elementales, ya que la estructura de la realidad no consiste sólo en ingredientes reduccionistas, tales como espacio, tiempo y partículas subatómicas, sino también, por ejemplo, en vida, pensamiento y calculabilidad.

Las *cuatro vías principales* de explicación que podrían constituir la primera teoría total de la realidad son:

Física cuántica: capítulos 2, 9, 11, 12, 13 y 14
Epistemología: capítulos 3, 4, 7, 10, 13 y 14
Teoría de la calculabilidad: capítulos 5, 6, 9, 10, 13 y 14
Teoría de la evolución: capítulos 8, 13 y 14

El próximo capítulo trata de la primera y más importante de esas cuatro vías: la física cuántica.

2. SOMBRAS

No existe puerta mejor, ni más abierta, para entrar en el estudio de la filosofía natural, que la consideración de los fenómenos físicos que se dan en una vela.

Michael Faraday

En sus populares disertaciones sobre ciencia en la Royal Institution, Michael Faraday solía instar a sus oyentes a aprender sobre el mundo mediante la observación de lo que ocurre cuando arde una vela. En su lugar, consideraré una linterna eléctrica. Ello resulta muy adecuado, ya que gran parte de la tecnología que utiliza se basa en los descubrimientos de Faraday.

Voy a describir algunos experimentos que demuestran fenómenos que constituyen el núcleo de la física cuántica. Experimentos de esta índole, con múltiples variaciones y perfeccionamientos, han constituido el pan nuestro de cada día en óptica cuántica durante muchos años. Si bien sus resultados son incontrovertibles, algunos de ellos siguen resultando difíciles de creer. Los experimentos básicos son notablemente sencillos y no requieren instrumentos especializados ni grandes conocimientos de matemáticas o física. En esencia, consisten, simplemente, en proyectar sombras. Sin embargo, las configuraciones de luz y sombra que una linterna común puede producir son muy extrañas, y, si se reflexiona sobre ellas, tienen extraordinarias ramificaciones. Su explicación requiere no sólo nuevas leyes físicas, sino un nuevo *nivel* de descripción y explicación que va más allá de lo que previamente se consideraba como ámbito de la ciencia. En primer lugar, muestran la existencia de universos paralelos. ¿Cómo? ¿Qué patrón concebible de sombras podría tener semejantes implicaciones?

Imaginemos que encendemos una linterna en una habitación oscura. La luz emana desde el filamento de la bombilla de la linterna y llena parte de un cono.

Para no complicar el experimento con luz reflejada, consideraremos las paredes de la habitación totalmente absorbentes, de color negro mate. También, y ya que sólo estamos realizando este experimento con la mente, podríamos imaginarnos una habitación de dimensiones astronómicas, de modo que no hubiese tiempo suficiente para que la luz alcanzase las paredes y volviese antes de haber acabado el experimento. La figura 2.1 ilustra la situación, pero resulta, en cierto modo, engañosa, ya que si estuviésemos observando la linterna desde uno de sus lados, no podríamos verla, y tampoco la luz que emite. La invisibilidad es una de las propiedades más evidentes de la luz: sólo la vemos cuando entra en nuestros ojos (cuando decimos que vemos algo, se trata, en realidad, del último objeto situado en nuestra línea de visión que ha sido afectado por la luz que ha entrado en ellos). No podemos ver la luz que pasa a nuestro alrededor. Si existiese algún objeto reflectante en el camino del rayo luminoso, o incluso un poco de polvo o unas gotas de agua que esparciesen la luz, podríamos ver dónde está, pero en el caso que nos ocupa no hay nada en el camino del rayo, y, puesto que lo observamos de lado, su luz no nos alcanza. Una representación ajustada de la realidad de lo que veríamos sería una imagen completamente negra. Si existiese una segunda fuente de luz, podríamos ver la linterna, pero seguiríamos sin ver su luz. Los rayos de luz, incluso de la más intensa que somos capaces de producir (por medio de láseres), se atraviesan unos a otros como si no existiesen en realidad.

La figura 2.1 muestra que la luz es más brillante cerca de la linterna y se vuelve más y más tenue a medida que el rayo se ensancha e ilumina una zona cada vez mayor. Un observador situado dentro del haz de luz que retrocediera lentamente desde la linterna, vería el reflector cada vez más pequeño y más tenue, hasta quedar reducido aun puntito. ¿Seguro? ¿Es posible que la luz se esparza cada vez más y sea cada vez más tenue de un modo ilimitado? La respuesta es que no. A unos diez mil kilómetros de la linterna, su luz resultaría demasiado tenue para ser detectada por el ojo humano y el observador no percibiría nada. Mejor dicho, un observador *humano* no percibiría nada; pero ¿le ocurriría lo mismo a un animal dotado de una visión más sensible?

Los ojos de las ranas son bastante más sensibles que los humanos, lo suficiente para introducir una diferencia significativa en nuestro experimento. Si el observador fuese una rana que se alejase de la linterna, nunca la perdería completamente de vista. Pero llegaría un momento en que la rana vería que la linterna empezaba a parpadear. Este parpadeo se produciría a intervalos regulares, que se harían más y más largos a medida que la rana se alejase, pero la brillantez de los destellos no disminuiría. A cien millones de kilómetros de la linterna, la rana percibiría un promedio de un destello al día, pero este destello sería tan potente como el observado a cualquier otra distancia.

Las ranas no nos pueden contar lo que ven, de modo que en los experimentos reales utilizamos fotomultiplicadores (detectores de luz aún más sensibles que los ojos de las ranas) y debilitamos la luz haciéndola pasar a través de filtros oscuros, puesto que no podemos observarla a cien mil kilómetros de distancia, pero el principio sigue siendo el mismo, al igual que el resultado: ni aparente oscuridad ni disminución uniforme, sino parpadeo, con destellos igualmente brillantes cualquiera que sea el filtro empleado. Este destello indica que existe un límite para la diseminación de la luz. Tomando prestada la terminología de los orfebres, podríamos decir que la luz no es infinitamente «maleable». Como el oro, una pequeña cantidad de luz puede diseminarse de modo uniforme sobre una gran área, pero a medida que se disemina llega un momento en que se vuelve grumosa. Incluso si fuese posible impedir que los átomos de oro tendieran a unirse, habría un punto más allá del cual no podrían ser subdivididos sin dejar de ser oro. Por lo tanto, el único modo de hacer aún más fina una hoja de oro del grueso de un átomo sería separar más y más sus átomos, dejando un creciente espacio entre ellos. Se llegaría así a un punto en el que ya no se podría hablar de una hoja continua. Si cada átomo estuviese separado de sus vecinos varios centímetros, por ejemplo, podríamos pasar la mano a través de la «hoja» sin tocar el oro. Del mismo modo, existe una última partícula o «átomo» de luz: el *fotón*.

Cada destello de luz percibido por la rana es un fotón que incide en su retina. Lo que sucede realmente cuando un rayo de luz se atenúa no es que los fotones se debiliten, sino que se distancian y aumenta el espacio vacío entre ellos (figura 2.2). Cuando un rayo es muy débil, puede resultar engañoso denominarlo «rayo», puesto que ya no es continuo. En los períodos en que la rana no ve nada, no es que la luz que llega a sus ojos sea demasiado débil para estimular su retina, sino que no reciben ningún estímulo luminoso.

Esta propiedad de presentarse sólo en «paquetes» de dimensiones discretas se denomina *cuantificación*. Un «paquete» individual —como un fotón— se llama *cuanto*. La teoría cuántica recibe su nombre de esta propiedad, que atribuye a toda magnitud física mensurable y no sólo a aquellas cosas, como la cantidad de luz o la masa de oro, que son cuantificadas porque las entidades concernidas, aunque aparentemente continuas, están en realidad formadas por partículas.

Incluso para magnitudes como la distancia (por ejemplo, entre dos átomos), la idea de un intervalo continuo de valores posibles es una idealización. En física no hay cantidades mensurables continuas. Hay infinidad de fenómenos sorprendentes en física cuántica, y comparada con ellos, como veremos, la cuantificación es uno de los menos espectaculares. No obstante, en cierto sentido sigue siendo la clave para los demás, puesto que, si todo es cuantificado, ¿cómo cambia una magnitud de un valor a otro? ¿Cómo va un objeto de un *lugar* a otro, si no existe un intervalo de lugares intermedios por los que pasa mientras está de camino? Lo explicaré en el capítulo 9, así que dejaremos de lado esta cuestión por el momento y volveremos a las inmediaciones de la linterna, donde el rayo nos parece continuo porque cada segundo emite unos 10^{14} (cien billones) de fotones hacia el ojo del observador.

¿Está la frontera entre luz y sombra netamente definida, o existe una zona gris? Normalmente, hay una zona gris bastante amplia, y una de sus causas se muestra en la figura 2.3. Existe una zona oscura (la *sombra*) allí donde no puede llegar la luz que emite el filamento. Hay también una zona iluminada, adonde llega la luz desde cualquier parte del filamento. Y, como éste no es un punto geométrico, sino que tiene un tamaño determinado, existe una zona intermedia entre la luz y la sombra, adonde sólo llega la luz de algunas partes del filamento (la *penumbra*). Un observador situado dentro de la zona de penumbra vería únicamente parte del filamento y notaría que la iluminación era allí menos intensa que en la zona que recibe toda la intensidad de la luz.

Sin embargo, el tamaño del filamento no es la única razón por la que las linternas reales proyectan zonas de penumbra. La luz se ve afectada por muchísimas circunstancias: el reflector situado tras la bombilla, el cristal que hay frente a ella, grietas e imperfecciones, etcétera. De modo que podemos esperar de una linterna real una configuración realmente complicada de luz y sombra a causa de su propia complejidad. Pero las propiedades incidentales de las linternas no constituyen el objeto de estos experimentos. Tras la cuestión de la luz de la linterna hay otra, más fundamental sobre la luz en general: ¿existe, en principio, algún límite para la nitidez de la sombra (en otras palabras, para la mayor o menor anchura de la penumbra)? Si, por ejemplo, la linterna estuviese construida con un material perfectamente negro (por tanto, no reflectante) y utilizásemos filamentos cada vez más pequeños, ¿podríamos reducir la zona de penumbra de un modo ilimitado?

La figura 2.3 induce a pensar que sí. Si el filamento careciese de dimensión, no habría penumbra. Pero al dibujarla he dado por sentado que la luz viaja únicamente en línea recta. La experiencia cotidiana nos indica que así es, ya que no podemos ver al otro de los ángulos, pero complejos experimentos demuestran que no siempre es cierto. En determinadas circunstancias, la luz se curva.

Esto resulta difícil de demostrar sólo con una linterna, ya que no es fácil conseguir filamentos minúsculos y superficies muy negras. Estas dificultades de orden práctico obstaculizan la plena comprensión de los límites que la física fundamental impone a la nitidez de las sombras. Afortunadamente, la curvatura de la luz puede ser también demostrada de otro modo. Supongamos que la luz de nuestra linterna pasa por dos pequeños orificios practicados en sendas pantallas opacas consecutivas —como ilustra la figura 2.4— y que la luz emergente es proyectada sobre una tercera pantalla. La cuestión es: si este experimento se repite con orificios cada vez más pequeños y aumentando progresivamente la distancia entre las dos primeras pantallas, ¿podremos acercar cada vez más, de modo ilimitado, la sombra (la zona de oscuridad total) a la línea recta que pasa por el centro de los dos orificios? ¿Puede la región iluminada entre la segunda y la tercera pantallas quedar confinada en un cono arbitrariamente estrecho? En la terminología de los orífices, estamos preguntándonos «cuán *dúctil* es la luz», cuán fino podemos llegar a hacer uno de sus rayos. El oro puede ser reducido a hebras de una diezmilésima de milímetro de espesor.

¡Resulta que la luz no es tan dúctil como el oro! Mucho antes de que los orificios se reduzcan a un diámetro de una diezmilésima de milímetro —de hecho, con orificios «anchísimos», de aproximadamente un milímetro de diámetro—, la luz empieza a rebelarse de modo notable. En lugar de pasar por ellos en línea recta, se resiste al confinamiento y se esparce después de cada orificio; y, al esparcirse, se «deshilacha». Cuanto más pequeño es el orificio, más se esparce la luz y se desvía del camino recto. Aparecen intrincadas conformaciones de luz y sombra. Sobre la tercera pantalla ya no vemos únicamente una zona iluminada y otra oscura, separadas por una penumbra, sino círculos concéntricos de anchura y brillo variables. Y también hay color, puesto que la luz blanca consiste en una mezcla de fotones de distintos colores. Cada color se esparce y «deshilacha» siguiendo

conformaciones ligeramente distintas. La figura 2.5 muestra una conformación típica de los colores producidos sobre la tercera pantalla por una luz blanca, tras pasar por los orificios de las dos primeras. Recordemos que lo único que estamos viendo es la proyección de una sombra. La figura 2.5 no es más que la sombra que proyectaría la segunda pantalla de la figura 2.4. Si la luz viajase únicamente en línea recta, no veríamos más que un minúsculo punto blanco (mucho más pequeño que el manchón brillante central de la figura 2.5), rodeado por una penumbra muy estrecha. El resto sería pura sombra, oscuridad total.

Por sorprendente que pueda resultar el que los rayos de luz se curven tras atravesar pequeños orificios, no se trata de algo, en mi opinión, fundamentalmente desconcertante. En todo caso, lo que nos importa para nuestro propósito es que lo hacen, lo cual implica que las sombras no tienen por qué parecer necesariamente siluetas de los objetos que las proyectan. Y, lo que es más, este fenómeno no es, simplemente, un emborronamiento debido a la penumbra. En efecto, un obstáculo con una intrincada serie de orificios puede proyectar una sombra de una conformación enteramente distinta.

La figura 2.6 muestra, aproximadamente en tamaño real, parte de la conformación de sombras proyectada a tres metros por dos ranuras rectilíneas paralelas practicadas en una barrera opaca. Las ranuras están distanciadas dos décimas de milímetro e iluminadas por un haz de lados paralelos de luz roja pura procedente de un láser situado al otro lado de la barrera. ¿Por qué luz de un láser y no de una linterna? Sólo porque la forma precisa de una sombra depende también del color de la luz que la proyecta; la luz blanca, como la producida por una linterna, contiene una mezcla de todos los colores visibles, de modo que puede proyectar sombras con bordes multicolores. Por consiguiente, en los experimentos sobre la forma precisa de las sombras resulta aconsejable la utilización de luz de un solo color. Podríamos colocar un filtro de color (un cristal coloreado, por

ejemplo) delante de la linterna de modo que sólo pudiese pasar la luz de dicho color. Sería una mejora, pero los filtros no son lo suficientemente selectivos. La luz de un láser constituye un método mejor, ya que puede ser regulada de manera muy exacta a fin de que emita sólo luz del color que deseamos, prácticamente sin la presencia de ningún otro.

Si la luz viajara en línea recta, el patrón de la figura 2.6 consistiría, simplemente, en un par de brillantes bandas separadas dos décimas de milímetro (espacio demasiado reducido para poder ser apreciado a simple vista a esta escala), con bordes netos y el resto de la pantalla en sombra. Pero, en realidad, la luz se curva de tal modo que produce varias bandas brillantes y oscuras y ningún borde neto. Si separamos las ranuras manteniéndolas dentro del haz del láser, la conformación de las bandas se desplazará en la misma proporción. En este aspecto, se comporta exactamente igual que una sombra corriente de gran escala. Ahora bien, ¿qué ocurre si intercalamos otras dos ranuras equidistantes, de modo que tengamos cuatro ranuras rectilíneas paralelas, a intervalos de una décima de milímetro? Cabría esperar que la conformación se pareciese casi exactamente a la de la figura 2.6. Después de todo, el primer par de ranuras proyecta por sí mismo las sombras de dicha figura y, como he dicho, el segundo par, por sí mismo, proyectaría la misma conformación, sólo que desplazada una décima de milímetro hacia un lado, casi sobre el mismo sitio. Sabemos también que los haces de luz se cruzan sin afectarse. De modo que los dos pares de ranuras juntos deberían, en principio, producir esencialmente la misma conformación, aunque con un brillo doble y algo más borrosa.

En realidad, sin embargo, no ocurre nada de eso. La forma real de la sombra de una barrera con cuatro ranuras rectilíneas paralelas, a intervalos de una décima de milímetro, se muestra en la figura 2.7 *a)*. A efectos de comparación, he repetido bajo ella, en la imagen *b)*, la correspondiente a las dos ranuras rectilíneas

paralelas a dos décimas de milímetro de separación. Es evidente que la sombra de las cuatro ranuras no es la combinación ligeramente desplazada proyectada por dos juegos de ranuras paralelas, sino que presenta una conformación nueva y más intrincada. En ella existen lugares, como el señalado con una X, oscuros en la conformación de cuatro ranuras, pero iluminados en la de dos. Estos lugares estaban iluminados cuando había sólo dos ranuras en la barrera, pero *pasaron a ser oscuros* cuando practicamos un segundo par de ranuras para el paso de la luz. La apertura de estas nuevas ranuras ha *interferido* la luz que llegaba previamente a X.

Así pues, añadir dos fuentes de luz provoca el oscurecimiento del punto X, mientras que eliminarlas lo ilumina de nuevo. ¿Cuál es la causa? Una posible explicación es que dos fotones se dirijan hacia X y reboten el uno en el otro como bolas de billar. Solo, cada fotón llegaría a X, pero al juntarse se interfieren y van a parar a otra parte. Demostraré en breve que esta explicación no puede ser cierta. Sin embargo, la idea básica que la inspira es incontrovertible: *algo* debe proceder del nuevo par de ranuras que impide que la luz del primero alcance el punto X. Pero ¿qué es? Lo podemos averiguar con la ayuda de algunos experimentos más.

En primer lugar, la conformación proyectada por las cuatro ranuras de la figura 2.7 *a)* aparece únicamente si todas ellas están iluminadas por el láser. Si sólo dos lo están, se proyecta la correspondiente a dos ranuras. Si se iluminan tres, aparece una nueva configuración, distinta de las anteriores. De modo que lo que causa la interferencia tiene que estar en el haz de luz. La conformación de dos ranuras aparece de nuevo si obstruimos dos de las cuatro con algo opaco, pero no si colocamos algo transparente. Dicho de otro modo, lo que interfiere queda anulado por cualquier cosa capaz de obstruir el paso de la luz, incluso algo tan insustancial como la niebla. Sin embargo, puede penetrar cualquier cosa que deje pasar la luz, incluso algo tan impenetrable (a la materia) como el diamante. Aunque introduzcamos en el

aparato complicados sistemas de espejos y lentes, mientras la luz pueda viajar desde las ranuras a cualquier punto de la pantalla, lo que observaremos en éste será una parte de la conformación que corresponde a las cuatro ranuras. Si luz proveniente de dos ranuras incide en algún punto de la pantalla, lo que veremos allí será parte de la conformación correspondiente a dos ranuras, y así sucesivamente.

Por consiguiente, lo que causa la interferencia se comporta como la luz. Se encuentra en todas partes en el haz de luz y en ningún lugar fuera de él. Es reflejado, transmitido o bloqueado por todo aquello capaz de reflejar, transmitir o bloquear la luz. Se estarán preguntando por qué insisto tanto en este punto. Es, sin duda, obvio que se trata de *luz*; es decir, lo que interfiere los fotones que salen de cada ranura son los que salen de las demás. Pero quizás se inclinarán a dudar de lo evidente tras el próximo experimento, el que culmina la serie.

¿Qué cabe esperar si realizamos estos experimentos con un *solo fotón cada vez*? Supongamos, por ejemplo, que colocamos la linterna tan lejos que sólo un fotón al día alcanza la pantalla. ¿Qué percibiría la rana si observara la luz desde la pantalla? Si es cierto que lo que interfiere los fotones son otros fotones, ¿no debería esta interferencia reducirse al disminuir el número de fotones? ¿No debería cesar completamente cuando circula un solo fotón por el aparato? Cabría esperar penumbras, ya que el fotón podría cambiar de trayectoria si chocara tangencialmente en el borde de una ranura. Pero, sin duda, lo que no podríamos observar nunca en la pantalla sería un lugar como X, que recibe fotones cuando están abiertas dos ranuras, pero *se queda a oscuras* al abrir dos más.

Sin embargo, esto es exactamente lo que observamos. Por más escasos que sean los fotones, la conformación de sombras sigue siendo la misma. Incluso cuando se realizan experimentos sucesivos con un solo fotón, éste no llega jamás a X con las cuatro ranuras abiertas, mientras que se reanuda el parpadeo sobre X si cerramos dos de ellas.

¿Podría ocurrir que el fotón se dividiese en fragmentos que, tras pasar por las ranuras, cambiasen de dirección y se reagrupasen? Podemos también descartar esta posibilidad. Si, una vez más, hacemos circular un fotón por el aparato, pero esta vez utilizando cuatro detectores, uno para cada ranura, en todos los casos sólo uno de ellos registra algo. Puesto que en dicho

experimento nunca observamos que dos detectores se activen a la vez, podemos afirmar que lo que detectan no se fracciona.

De modo que, si los fotones no se dividen en fragmentos y no son desviados por otros fotones, ¿qué los desvía? Cuando un solo fotón circula por el aparato, ¿qué puede salir de las otras ranuras capaz de interferido?

Recapitulemos. Hemos encontrado que cuando un fotón circula por el aparato:

—pasa por una de las ranuras, luego algo lo interfiere y lo desvía de un modo que depende de cuáles de las demás ranuras estén abiertas;

—lo que interfiere al fotón ha pasado por algunas de las otras ranuras;

—lo que interfiere al fotón se comporta exactamente como un fotón...

—... sólo que no puede ser visto.

A partir de ahora denominaré «fotones» a lo que causa las interferencias. Es lo que son. Así pues, ahora parece haber dos clases de fotones, a los que llamaré, temporalmente, fotones *tangibles* y fotones *fantasmas*. Los fotones tangibles son los que podemos ver o detectar con instrumentos, mientras que los fantasmas son intangibles (invisibles) y sólo es posible detectarlos indirectamente por sus efectos de interferencia sobre aquéllos. (Veremos más adelante que no existe diferencia intrínseca entre unos y otros: cada fotón es tangible en un universo y fantasma en todos los demás universos paralelos. Pero más vale que avance paso a paso en mi exposición.) Lo único que hemos inferido hasta aquí es que cada fotón tangible va acompañado por un séquito de fotones fantasmas, y que mientras un fotón tangible pasa por una de las cuatro ranuras un número variable de fotones fantasmas lo hace por las otras tres. Puesto que aparecen distintas conformaciones de interferencias si abrimos ranuras en otros lugares de la barrera —siempre que estén dentro del haz de luz—, cada vez que un fotón tangible llega a la parte iluminada de la barrera, debe ir acompañado por un número variable de fotones fantasmas. Por lo tanto, hay más fotones fantasmas que tangibles. ¿Cuántos? La experimentación no nos puede concretar su número máximo, pero sí un mínimo aproximado. En el laboratorio, la superficie máxima que podemos iluminar convenientemente es de un metro cuadrado, mientras que la dimensión mínima manejable para el diámetro de un orificio es de una milésima de milímetro,

así que tenemos aproximadamente 10^{12} (un billón) de posibles posiciones para un orificio de una milésima de milímetro de diámetro en una barrera de un metro cuadrado de superficie. Por lo tanto, un mínimo de un billón de fotones fantasmas acompaña a cada fotón tangible.

Hemos deducido, pues, la existencia de un mundo prodigiosamente numeroso, complicado y oculto, de fotones fantasmas. Viajan a la velocidad de la luz, rebotan en los espejos, son refractados por las lentes y son detenidos por las barreras opacas y los filtros del color adecuado. Sin embargo, no activan los detectores más sensibles. En el universo, lo único que permite observar la existencia de los fotones fantasmas son las alteraciones que causan en la trayectoria de los fotones tangibles mediante el fenómeno de la interferencia. Los fotones fantasmas pasarían completamente inadvertidos de no ser por este fenómeno y por las sorprendentes conformaciones de sombras que provoca.

La interferencia no es una propiedad exclusiva de los fotones. La teoría cuántica predice, y lo confirma la experimentación, que se da en toda clase de partículas. Deben existir, pues, infinidad de neutrones fantasmas que acompañan a cada neutrón tangible, miríadas de electrones fantasmas por cada electrón tangible, y así sucesivamente. Las partículas fantasmas sólo son detectables indirectamente, por su interferencia con el movimiento de sus homólogas tangibles.

Se deduce, pues, que la realidad es mucho mayor de lo que parece, y que en su mayor parte es invisible. Los objetos y sucesos que nuestros sentidos y nuestros instrumentos pueden observar de manera directa no son más que la punta del iceberg.

Ahora bien, las partículas tangibles tiene una propiedad que nos autoriza a denominarlas, colectivamente, *universo*. No es otra que su propiedad definitoria de ser tangibles, es decir, de interactuar unas con otras y, por lo tanto, ser directamente detectables por instrumentos y órganos sensoriales formados por otras partículas tangibles. A causa de los fenómenos de interferencia, no están *completamente* aisladas del resto de la realidad (es decir, de las partículas fantasmas). Si lo estuviesen, nunca habríamos descubierto que la realidad va mucho más allá de las partículas tangibles. Las partículas tangibles forman el universo que observamos a nuestro alrededor en nuestra vida cotidiana, así como el universo descrito por la física clásica o precuántica.

Por razones similares, podríamos pensar en denominar a las

partículas fantasmas, colectivamente, *universo paralelo*, puesto que también son afectadas por las partículas tangibles únicamente a través de los fenómenos de interferencia. Pero podemos hacer algo mejor. Y es que resulta que las partículas fantasmas están separadas entre sí del mismo modo que el universo de las partículas tangibles lo está de ellas. Dicho de otro modo, no forman un único y homogéneo universo paralelo, infinitamente mayor que el tangible, sino, más bien, un enorme número de universos paralelos, cada uno de los cuales se asemeja en su composición al tangible y obedece a las mismas leyes de la física, pero que se diferencian porque sus partículas están en distintas posiciones en cada universo.

Una observación sobre la terminología. La palabra «universo» se ha usado tradicionalmente para significar «toda la realidad física». En este sentido, sólo puede existir un universo. Podríamos mantener esta definición y decir que la entidad que estamos acostumbrados a definir como «el universo» —es decir, toda la materia y la energía directamente perceptibles que nos rodean, así como el espacio— no es más que una mínima parte del verdadero universo. Deberíamos entonces inventar un nuevo nombre para esta pequeña porción tangible. Pero la mayoría de los físicos prefieren seguir utilizando la palabra «universo» para denominar a la entidad de siempre, aun cuando ésta resulte ser ahora sólo una pequeña porción de la realidad física. Un nuevo término, *multiverso*, ha sido acuñado para denominar la totalidad de la realidad física.

Los experimentos de interferencia con una sola partícula, como los que he descrito, nos demuestran que el multiverso existe y que contiene múltiples partículas fantasmas que son la contrapartida de las partículas del universo tangible. Para ir aún más allá, y llegar a la conclusión de que el multiverso está dividido en una especie de universos paralelos, debemos considerar los fenómenos de interferencia que afectan a más de una partícula tangible. El modo más simple de hacerlo consiste en preguntarnos, a modo de «experimento mental», qué debe de ocurrir a nivel microscópico cuando los fotones fantasmas chocan con un objeto opaco. Son detenidos, desde luego. Lo sabemos porque la interferencia cesa cuando bloqueamos el paso de los fotones fantasmas colocando un obstáculo opaco. Pero ¿por qué? ¿Qué los detiene? Podemos descartar la respuesta que parece más evidente: que son absorbidos —como lo serían los fotones

tangibles— por los átomos tangibles de la barrera. En primer lugar, porque sabemos que los fotones fantasmas no interactúan con átomos tangibles. Y, en segundo lugar, porque podemos verificar, mediante la medición de los átomos en la barrera (o, más exactamente, reemplazándola por un detector), que éstos ni absorben energía ni cambian de estado en lo más mínimo a menos que reciban el impacto de fotones tangibles. Los fotones fantasmas carecen, pues, de efecto sobre ellos.

En otras palabras, los fotones fantasmas y los tangibles se ven afectados por igual cuando alcanzan una determinada barrera, pero ésta no se ve afectada del mismo modo por las dos clases de fotones. De hecho, de acuerdo con nuestros conocimientos actuales, la barrera no se ve afectada en lo más mínimo por los fotones fantasmas. Ésta es, ciertamente, la propiedad definitoria de los fotones fantasmas, ya que si algún material fuese afectado por ellos de un modo susceptible de ser observado, dicho material podría ser utilizado como detector de fotones fantasmas y todo el fenómeno de sombras e interferencia sería distinto de como lo he descrito.

Existe, por lo tanto, alguna clase de barrera fantasma en el mismo lugar que la barrera tangible. No hace falta un gran esfuerzo de imaginación para concluir que dicha barrera no puede estar constituida más que por *átomos fantasmas*, que sabemos que deben existir como contrapartida de los átomos tangibles en la barrera. Hay presentes muchos de ellos por cada átomo tangible. En realidad, la densidad total de átomos fantasmas en la niebla más ligera es tal, que bastaría para detener a un carro de combate, y no digamos a un fotón, *si pudiesen afectarlos*. Del hecho de que, como hemos visto, las barreras parcialmente transparentes tengan el mismo grado de transparencia para los fotones fantasmas que para los tangibles se sigue que no todos los átomos fantasmas que se interponen en el camino de un determinado fotón fantasma se ven implicados a la hora de bloquearle el paso. Cada fotón fantasma se encuentra con una clase de barrera muy parecida a aquella con la que se enfrenta su contrapartida tangible, la cual consiste únicamente en una pequeña proporción de todos los átomos fantasmas presentes.

Por la misma razón, cada átomo fantasma de la barrera sólo puede interactuar con una pequeña proporción de los átomos fantasmas que se encuentran cerca de él, con los que forma una barrera muy parecida a la tangible, y así sucesivamente. Toda la

materia, y todos los procesos físicos, tienen esta estructura. Si la barrera tangible es la retina de la rana, allí debe haber múltiples retinas fantasmas, cada una de ellas capaz de detener únicamente una de las contrapartidas de cada fotón. Cada retina fantasma interactúa tan sólo con los correspondientes fotones fantasmas y la correspondiente rana fantasma, y así sucesivamente. En otras palabras, las partículas están agrupadas en universos paralelos. Son «paralelos» en el sentido de que dentro de cada universo las partículas interactúan entre sí como en el universo tangible, mientras que cada universo sólo afecta a los demás ligeramente, mediante los fenómenos de interferencia.

Hemos llegado así a la conclusión de la cadena de razonamientos que empieza con sombras de extrañas conformaciones y acaba con universos paralelos. Cada paso nos adentra en la comprobación de que el comportamiento de los objetos que observamos sólo puede ser explicado por la existencia de objetos presentes no observables, dotados de determinadas propiedades. El núcleo del argumento es que los fenómenos de interferencia de partículas aisladas descartan inequívocamente la posibilidad de que el universo tangible que nos rodea sea todo cuanto existe. No hay discusión posible sobre el hecho de que dichos fenómenos de interferencia ocurren, y, sin embargo, la existencia del multiverso es aún un punto de vista minoritario entre los físicos. ¿Por qué?

La respuesta, lamento decirlo, no habla mucho en favor de la mayoría. Hablaré más extensamente acerca de este tema en el capítulo 13, pero, de momento, permítaseme señalar que los argumentos que he expuesto en el presente capítulo son convincentes sólo para quienes buscan explicaciones. Aquellos que se contentan con meras predicciones y carecen de un fuerte deseo de comprender cómo llegan a producirse los resultados predichos para los experimentos pueden, si lo desean, negar, simplemente, la existencia de todo lo que no sean las entidades que hemos denominado «tangibles». Hay quienes, como los instrumentalistas y los positivistas, adoptan esta actitud por fidelidad a unos principios filosóficos. He expuesto ya lo que pienso de esos principios y por qué. Otros, simplemente, no quieren pensar en ello. Después de todo, es una conclusión tan *amplia*, que puede resultar turbadora al principio. Creo, sin embargo, que cometen un error. Espero tener éxito y convencer a los lectores que me acompañen de que la comprensión del multiverso es condición

previa para comprender la realidad del mejor modo posible. No lo digo imbuido de un espíritu de sombría determinación de buscar la verdad por más desagradable que sea (si bien creo que ésta sería la actitud que adoptaría si no tuviera más remedio), sino, muy al contrario, porque la concepción del mundo resultante está mucho más integrada y es mucho más sensata, en infinidad de aspectos, que cualquiera de las precedentes y, ciertamente, que el cínico pragmatismo que demasiado a menudo sirve hoy como sustituto de una concepción del mundo para ciertos científicos.

«¿Por qué no podemos decir, simplemente», preguntan algunos físicos pragmáticos, «que los fotones se comportan *como* si interactuasen con entidades invisibles? ¿Por qué no dejamos las cosas así? ¿Por qué debemos tomar partido acerca de si esas entidades existen realmente?» Una variante más exótica de lo que es, en esencia, la misma idea es la siguiente: «Un fotón tangible es real; un fotón fantasma no es más que una forma en que el fotón real hubiese podido actuar, pero no lo hizo. Por consiguiente, la teoría cuántica trata de la *interacción de lo real con lo posible*.» Esto último, al menos, suena razonablemente profundo. Pero, por desgracia, quienes adoptan cualquiera de esas dos posturas —incluyendo algunos científicos que deberían estar mejor informados—, derivan invariablemente hacia la palabrería vana a partir de esas tomas de posición. Mantengamos, pues, la calma. El factor clave no es otro que el hecho de que un fotón real, tangible, *se comporta de diversas maneras* en función de qué caminos estén abiertos en el aparato para que pueda circular por ellos algo que lo interfiera. Algo circula por esos caminos, y resistirse a considerarlo «real» es un mero juego de palabras. «Lo posible» no puede interactuar con lo real: las entidades inexistentes no pueden desviar a las reales de su camino. Si un fotón es desviado, debe serlo por algo, algo que yo denomino «fotón fantasma». Atribuirle un nombre no lo convierte en real, pero de ningún modo puede ser cierto que un hecho real, tal como la llegada y detección de un fotón real, sea causado por un suceso imaginario, por algo que dicho fotón «hubiese podido hacer», pero no hizo. Sólo lo que sucede realmente puede hacer que otras cosas sucedan realmente. Si el complejo movimiento de los fotones fantasmas en un experimento de interferencia fuese una mera posibilidad que no se ha hecho realidad, los fenómenos de interferencia que observamos no habrían llegado a producirse realmente.

La razón de que los fenómenos de interferencia sean tan

débiles y difíciles de detectar la hallamos en las leyes cuántico-mecánicas que los rigen. Dos implicaciones de esas leyes, en particular, resultan relevantes. En primer lugar, cada partícula subatómica tiene contrapartidas en otros universos y es interferida únicamente por éstas. No se ve, pues, afectada por ninguna otra partícula de dichos universos. Por consiguiente, la interferencia sólo puede ser observada en situaciones concretas en que los caminos de una partícula y sus contrapartidas fantasmas se separan para converger después (como cuando un fotón y un fotón fantasma se dirigen hacia el mismo punto de la pantalla). Incluso la coincidencia temporal debe ser la adecuada: si en uno de los caminos se produce un retraso, la interferencia será de menor intensidad o no llegará a ocurrir. En segundo lugar, la detección de interferencias entre dos universos cualesquiera requiere que interactúen *todas las partículas cuyas posiciones y demás atributos no sean idénticos* en dichos universos. En la práctica, ello significa que sólo la interferencia entre universos muy parecidos es lo suficientemente fuerte para ser detectada. En los experimentos que he descrito, por ejemplo, los universos que se interfieren sólo difieren en la posición de un fotón. Si un fotón afecta en sus viajes a otras partículas, y, en particular, si es observado, tanto esas partículas como el observador serán diferentes en otros universos distintos. De ahí que las subsiguientes interferencias relacionadas con dicho fotón sean imposibles de detectar en la práctica, puesto que la necesaria interacción entre todas las partículas afectadas resulta demasiado compleja para que sea factible observarla. Debo señalar aquí que la frase habitual para describir este hecho, es decir, que «la observación destruye la interferencia», resulta extremadamente engañosa en tres sentidos. En primer lugar, porque sugiere alguna especie de efecto psicocinético del «observador» sobre fenómenos físicos elementales, efecto que, en realidad, no existe. En segundo lugar, porque la interferencia no es «destruida»: simplemente, resulta muy difícil de observar, puesto que hacerlo requiere controlar el comportamiento preciso de muchas más partículas. Y, en tercer lugar, porque no es únicamente la «observación», sino cualquier efecto del fotón sobre su entorno, lo que dependerá de qué camino haya tomado éste.

Para facilitar las cosas a aquellos lectores que conozcan otros trabajos sobre física cuántica, debo referirme brevemente al argumento que he presentado en este capítulo y al modo en que el

tema se plantea de modo habitual. Quizás porque el debate se inició entre físicos teóricos, su punto de partida tradicional ha sido la propia teoría cuántica. Se expone ésta del modo más cuidadoso posible, para tratar después de entender lo que nos dice sobre la realidad. Éste es el único planteamiento posible, si se desea comprender los más sutiles detalles de los fenómenos cuánticos. Pero resulta un planteamiento innecesariamente complicado en relación con la cuestión de si la realidad consiste en uno o varios universos. Por ello no lo he seguido en el presente capítulo. Ni siquiera he expuesto ninguno de los postulados de la teoría cuántica. He descrito, simplemente, algunos fenómenos físicos y sacado algunas conclusiones inevitables.

Pero si queremos empezar por la teoría, nos encontraremos con dos cosas sobre las que hay acuerdo unánime. La primera es que la teoría cuántica no tiene rival en cuanto a capacidad para predecir los resultados de los experimentos, incluso si utilizamos sus ecuaciones ciegamente sin preocuparnos demasiado por lo que significan. La segunda es que la teoría cuántica nos dice algo nuevo e insólito sobre la naturaleza de la realidad. La disputa no estriba sólo en qué es, exactamente, lo que nos dice. El físico Hugh Everett fue el primero en comprender con claridad (en 1957, unos treinta años después de que se convirtiese en la base de la física subatómica) que la teoría cuántica describe un multiverso. Desde entonces, la discusión sobre si la teoría admite alguna otra interpretación (o reinterpretación, reformulación, modificación, etcétera) que describa un único universo, sin dejar por ello de predecir correctamente los resultados experimentales, no ha cesado de agudizarse. Dicho de otro modo: ¿nos obliga la aceptación de las predicciones que son fruto de la teoría cuántica a admitir la existencia de universos paralelos?

Me parece que tanto esta cuestión como el tono dominante en el debate subsiguiente están mal enfocados. Ciertamente, es lo correcto, y lo que cabe esperar de físicos teóricos como yo es dedicar grandes esfuerzos a tratar de comprender la estructura formal de la teoría cuántica, pero no a expensas de perder de vista nuestro objetivo primordial, que no es otro que la comprensión de la realidad. Incluso si las predicciones de la teoría cuántica pudiesen, de algún modo, ser hechas refiriéndose a un solo universo, los fotones seguirían proyectando sombras del modo que

he descrito. Aun desconociéndolo todo sobre la teoría cuántica, resultaría evidente que esas sombras no pueden ser el resultado de ninguna única historia del fotón en su trayectoria desde la linterna hasta el ojo del observador. Resultan incompatibles con cualquier explicación únicamente en términos de los fotones que podemos ver. O únicamente en términos de la barrera que podemos ver. O únicamente en términos del universo que podemos ver. Por consiguiente, incluso si la mejor teoría de que dispusieran los físicos no hiciese mención de universos paralelos, ello sólo significaría que necesitábamos una teoría mejor, una teoría que se refiriese a esos universos que no vemos para explicar lo que vemos.

Así pues, ¿nos obliga la aceptación de las predicciones fruto de la teoría cuántica a admitir la existencia de universos paralelos? No necesariamente. Siempre podremos reinterpretar cualquier teoría en clave instrumentalista, de modo que no nos obligue a aceptar nada sobre la realidad. Pero eso no resolvería el problema. Como he dicho, no necesitamos profundas teorías que nos confirmen la existencia de universos paralelos: los fenómenos de interferencia entre partículas aisladas ya lo hacen. Las necesitamos para explicar y predecir dichos fenómenos: para que nos digan cómo son los demás universos, qué leyes obedecen, de qué modo se afectan unos a otros y cómo encaja todo ello con las bases teóricas de otras materias. Y esto es, precisamente, lo que hace la teoría cuántica. La teoría cuántica de los universos paralelos no es el problema, sino la solución. No se trata de una interpretación problemática y opcional, surgida de arcaicas consideraciones teóricas, sino de la explicación —la única sostenible— de una notable realidad que contradice cualquier intuición.

Hasta aquí he utilizado una terminología que sugiere que uno de los múltiples universos paralelos difiere de los demás por ser «tangible». Era algo temporal, y es tiempo ya de cortar esta última amarra con la concepción clásica, monouniversal, de la realidad. Volvamos a nuestra rana. Hemos visto ya que la historia de la rana que permanece sentada observando pacientemente la lejana linterna día tras día, a la espera de la llegada del destello cotidiano, no es toda la historia. Debe haber también ranas fantasmas en universos fantasmas, que coexistan con la tangible y esperen también la llegada de fotones.

Supongamos que la rana ha sido adiestrada para saltar cuando ve un destello. Al principio del experimento, la rana tangible tendrá gran cantidad de contrapartidas fantasmas, todas inicialmente iguales, pero que poco después ya no lo serán. Ninguna de ellas tiene demasiadas probabilidades de ver inmediatamente un fotón. Pero lo que en un universo concreto es un acontecimiento aislado, se convierte en un suceso común en el multiverso, entendido como un todo. En cada instante, en algún lugar del multiverso, habrá unos pocos universos en los que un fotón incidirá en las retinas de las ranas de dichos universos, y, cuando eso suceda, las ranas saltarán.

¿Qué es, exactamente, lo que hace saltar a las ranas? En su universo, obedecen las mismas leyes de la física que las ranas tangibles. Un fotón fantasma correspondiente a su universo ha incidido en su retina fantasma. Una de las moléculas fantasmas sensibles a la luz de su retina fantasma ha respondido y ha experimentado una serie de complejos cambios químicos, a los que el nervio óptico fantasma ha respondido a su vez. Éste ha transmitido un mensaje al cerebro fantasma de la rana, la cual ha experimentado, en consecuencia, la sensación de percibir un destello.

¿O debería decir la sensación *fantasma* de percibir un destello? No, sin la menor duda. Si los observadores «fantasmas»—sean éstos ranas o personas— son reales, sus sensaciones deben ser igualmente reales. Cuando observan lo que podríamos denominar un objeto fantasma, lo ven como tangible. Lo observan con los mismos medios y según la misma definición que aplicamos cuando decimos que el universo que observamos es «tangible». La tangibilidad es relativa para un determinado observador, de modo que, objetivamente, no hay dos clases de fotón, tangible y fantasma, ni dos clases de rana, ni de universo, uno tangible y el resto fantasma. No existe nada, en la descripción que he ofrecido acerca de la formación de sombras, o de los fenómenos relacionados, que implique distinción entre objetos «tangibles» y «fantasmas», aparte de la mera afirmación de que una de las copias es «tangible». Cuando introduje los conceptos de fotones tangibles y fotones fantasmas, parecía establecer distinciones entre ellos al afirmar que «nosotros» podemos ver los primeros, pero no los segundos.

Pero ¿quiénes somos «nosotros»? Mientras lo escribía, miríadas de David lo escribían a su vez. También ellos hacían una distinción entre fotones tangibles y fotones fantasmas, pero los fotones que para ellos son «fantasmas» incluyen los que para mí son «tangibles», mientras que los que ellos denominan «tangibles» están entre los que son «fantasmas» para mí.

Las copias de cualquier objeto no sólo carecen de posición alguna de privilegio en la explicación de las sombras que acabo de esbozar, sino de la más mínima preeminencia en la explicación matemática completa proporcionada por la teoría cuántica. Puedo pensar, subjetivamente, que me distingo de mis copias, y soy la única «tangible», porque que puedo percibirme de modo directo, y a ellas no, aunque debo hacerme a la idea de que todas ellas sienten lo mismo acerca de sí mismas.

Muchos de esos David están ahora mismo escribiendo estas palabras. Algunos expresan mejor estos pensamientos. Otros han ido por una taza de té.

TERMINOLOGÍA

Fotón. Partícula de luz.

Fotones tangibles y fotones fantasmas. Para facilitar la exposición, y sólo en este capítulo, denomino a las partículas de este universo *tangibles*, y a las de los demás, *fantasmas*.

Multiverso. Totalidad de la realidad física. Contiene múltiples universos.

Universos paralelos. Son «paralelos» en el sentido de que dentro de cada uno de ellos las partículas interactúan unas con otras del mismo modo que en el universo tangible, pero cada universo sólo afecta a los demás débilmente, mediante los fenómenos de interferencia.

Teoría cuántica. Teoría de la física del multiverso.

Cuantificación. Propiedad de poseer una serie de posibles valores discreta (más que continua). La teoría cuántica recibe este nombre porque enuncia que todas las cantidades mensurables poseen esta propiedad. Sin embargo, el efecto cuántico más notable no es la cuantificación, sino la interferencia.

Interferencia. Efecto de una partícula de un universo sobre su contrapartida de otro. La interferencia entre fotones es la responsable de que las sombras no se limiten a ser meras siluetas de los obstáculos que las producen y presenten mucho mayor complejidad.

SUMARIO

En los experimentos de interferencia se observa que hay lugares iluminados en una conformación de sombras que se oscurecen al practicar nuevas aberturas en la barrera que produce dichas sombras. Esto ocurre incluso cuando el experimento se realiza con partículas individuales.

Una cadena de razonamientos basada en este hecho descarta la posibilidad de que el universo que percibimos a nuestro alrededor constituya la totalidad de la realidad. De hecho, la totalidad de la realidad física, el multiverso, contiene gran número de universos paralelos.

La física cuántica constituye una de las cuatro vías mayores de explicación. La siguiente es la epistemología, *o teoría del conocimiento*.

3. RESOLUCIÓN DE PROBLEMAS

No sé qué es más sorprendente, si el propio comportamiento de las sombras o el hecho de que la observación de algunas conformaciones de luz y sombra nos obligue a revisar tan radicalmente nuestra concepción de la estructura de la realidad. El argumento que he esbozado en el capítulo anterior es, independientemente de su controvertible conclusión, un ejemplo típico de razonamiento científico. Vale la pena reflexionar sobre el carácter de este razonamiento, que constituye por sí mismo un fenómeno natural tan sorprendente y lleno de implicaciones, por lo menos, como la física de las sombras.

A aquellos que preferirían atribuir a la realidad una estructura más prosaica, tal vez les parezca desproporcionado —e incluso injusto— que consecuencias tan trascendentales puedan seguirse del hecho de que un diminuto punto de luz sobre una pantalla se encuentre *aquí*, en vez de estar *allí*. Pero así es, y no se trata de la primera vez en la historia de la ciencia que ocurre algo semejante, ni mucho menos. A este respecto, el descubrimiento de otros universos recuerda mucho el de los planetas hecho por los primeros astrónomos. Antes de que mandásemos sondas a la Luna y a algunos planetas, *toda* nuestra información sobre estos últimos procedía de manchas de luz (y de otras radiaciones) cuya presencia era observada en lugares donde no se esperaba encontrarla. Consideremos cómo se descubrió el hecho fundamental y definitorio acerca de los planetas: el de que no son estrellas. Si miramos el cielo por la noche durante unas horas, observaremos que las estrellas parecen evolucionar alrededor de un determinado punto del espacio. Lo hacen rígidamente, manteniendo posiciones fijas entre sí. Según la explicación tradicional, el cielo nocturno estaba constituido por una enorme «esfera celestial» que giraba alrededor de una Tierra fija, y las estrellas eran, o bien agujeros en dicha esfera, o bien cristales resplandecientes engarzados en ella. Sin embargo, entre los millares de puntos de luz visibles de noche al ojo desnudo, algunos de los más brillantes, cuando son observados a lo largo de períodos prolongados de tiempo, no se comportan de esa manera rígida, como si estuviesen fijos en la bóveda celestial, sino que parecen pasearse por el espacio con movimientos más complejos.

Son los denominados «planetas», palabra que deriva del griego *planétes*, que significa «vagabundo» o «errante». Su vagabundeo indicaba que la explicación de la esfera celestial resultaba inadecuada.

Las sucesivas explicaciones sobre el movimiento de los planetas han tenido un importante papel en la historia de la ciencia. La t*eoría heliocéntrica* de Copérnico situaba los planetas y la Tierra en órbitas circulares alrededor del Sol. Kepler descubrió que esas órbitas no eran circulares, sino elípticas. Newton explicó las elipses mediante su ley de la gravitación al cuadrado de sus distancias, la cual fue utilizada más tarde para predecir que esas fuerzas de atracción entre los planetas causarían pequeñas desviaciones en sus órbitas elípticas. La observación de estas desviaciones condujo al descubrimiento de un nuevo planeta —Neptuno— en 1846, uno de los muchos hallazgos que corroboraron de manera espectacular la teoría de Newton. Sin embargo, algunas décadas más tarde, la teoría general de la relatividad de Einstein nos proporcionaría una explicación fundamentalmente distinta de la gravedad, en términos de tiempo y espacio curvos, y predeciría asf trayectorias ligeramente distintas. Predijo correctamente, por ejemplo, que el planeta Mercurio se desvía cada año una diezmilésima de grado de la posición prevista por la teoría de Newton. Predijo también que la luz estelar, al pasar cerca del Sol, era desviada por la gravedad el doble de lo previsto por dicha teoría. La comprobación de esta desviación, efectuada por Arthur Eddington en 1919, es considerada a menudo el momento a partir del cual la visión newtoniana del mundo dejó de ser sostenible desde un punto de vista racional. (Irónicamente, las reevaluaciones modernas de la exactitud del experimento de Eddington sugieren que esta apreciación podría haber sido prematura.) Dicho experimento, que ha sido repetido con gran exactitud, incluía la medición de las posiciones de manchas (las imágenes de estrellas cercanas al limbo del Sol durante un eclipse) sobre una placa fotográfica.

A medida que las predicciones astronómicas se hacían más exactas, disminuían las diferencias en el aspecto del cielo nocturno predichas por las sucesivas teorías, y ha habido que construir telescopios e instrumentos de medida cada vez más potentes para detectar esas menguantes discrepancias. Sin embargo, las explicaciones en que se basan esas predicciones no han convergido, sino todo lo contrario: como ya he señalado, ha

habido una sucesión de cambios revolucionarios. Así pues, la observación de efectos físicos cada vez más pequeños ha provocado cambios cada vez más grandes en nuestra concepción del mundo. Se diría que sacamos conclusiones cada vez mayores de evidencias cada vez menores. ¿Qué es lo que justifica esas inferencias? ¿Podemos estar seguros de que sólo porque una estrella aparezca desplazada unos milímetros en la placa fotográfica de Eddington el espacio y el tiempo deben ser curvos, o de que sólo porque un fotodetector situado en una determinada posición no registre un «impacto» con luz débil, deben existir universos paralelos?

Sin duda, lo que acabo de decir subraya tanto la fragilidad como el carácter indirecto de toda evidencia experimental. No percibimos directamente las estrellas, ni las manchas en una placa fotográfica, ni cualquier otro objeto o acontecimiento externo. Sólo vemos las cosas cuando sus imágenes aparecen en nuestras retinas, y ni siquiera percibimos estas imágenes hasta que han dado lugar a impulsos eléctricos en nuestros nervios que, a su vez, han sido recibidos e interpretados por nuestros cerebros. De modo que la evidencia física que influye directamente en nosotros se mide en milésimas de milímetro (la distancia que separa las fibras nerviosas en el nervio óptico) y en centésimas de voltio (la carga del potencial eléctrico de nuestros nervios, que marca la diferencia entre percibir una cosa u otra).

No obstante, no otorgamos la misma importancia a todas nuestras impresiones sensoriales. En la experimentación científica, por ejemplo, hacemos grandes esfuerzos por acercar a nuestra percepción aquellos aspectos de la realidad exterior que creemos que nos pueden ayudar a evaluar las teorías rivales que estamos considerando. Antes incluso de realizar una observación, decidimos con todo cuidado qué, cuándo y dónde debemos buscar. A menudo utilizamos instrumentos complejos, diseñados específicamente, como telescopios y fotomultiplicadores. Pero, por más perfeccionados que estén los equipos empleados, y por más consistentes que sean las causas externas a las que atribuimos los resultados de sus observaciones, en última instancia los percibimos únicamente a través de nuestros órganos sensoriales. No hay escapatoria al hecho de que los seres humanos somos criaturas pequeñas, dotadas sólo de unos pocos e inadecuados canales para recibir toda la compleja información que nos llega del exterior. Interpretamos esta información como prueba de la

existencia de un universo exterior grande y complejo (o un multiverso), pero cuando sopesamos esas pruebas, no contemplamos, literalmente, nada más que unos débiles circuitos de corriente eléctrica que cosquillean nuestros cerebros.

¿Qué justifica las inferencias que extraemos de esos circuitos? No se trata, sin duda, de una cuestión de deducción lógica. No hay manera de *probar*, mediante estas o cualesquiera otras observaciones, no ya que el universo exterior, o multiverso, existe realmente, sino ni siquiera que las corrientes eléctricas recibidas por nuestros cerebros tengan alguna relación con él. Todo lo que percibimos puede ser una ilusión, un sueño. Las ilusiones y los sueños son, después de todo, cosa corriente. El *solipsismo* —teoría según la cual sólo existe una mente, y lo que aparenta ser la realidad externa no es más que un sueño desarrollado en esa mente— no puede ser rechazado lógicamente. La realidad *podría* consistir en una sola persona, presumiblemente usted, que soñara las experiencias de toda una vida. O podría consistir tan sólo en usted y yo. O en el planeta Tierra y sus habitantes. Y por más evidencias —de la clase que fuera— que soñáramos acerca de la existencia de otras personas, de otros planetas, de otros universos, ello no probaría ni la existencia real de esas cosas ni su número.

Puesto que el solipsismo, e infinidad de teorías relacionadas con él, son compatibles, desde un punto de vista lógico, con el hecho de que usted perciba cualquier evidencia observable posible, resulta que no puede deducir lógicamente nada acerca de la realidad a partir de dicha evidencia observable. ¿Cómo puedo, entonces, decir que el comportamiento observado de las sombras «refuta» la teoría de que existe un solo universo, o que la observación de los eclipses realizada por Eddington hace que la concepción newtoniana del mundo resulte «racionalmente insostenible»? ¿Cómo es posible? Si «refutar» no significa «demostrar la falsedad», ¿qué significa? ¿Por qué deberíamos sentirnos inclinados a cambiar nuestra concepción del mundo, o incluso cualquier opinión, sobre la base de que algo quede «refutado» en este sentido? Esta crítica parece proyectar dudas sobre la totalidad de la ciencia, es decir, sobre cualquier razonamiento acerca de la realidad exterior que se base en la evidencia observable.

Si el razonamiento científico no se concreta en secuencias de deducciones lógicas a partir de esa evidencia, ¿en qué se

concreta? ¿Por qué deberíamos aceptar sus conclusiones?

Esta cuestión se conoce como el «problema de la inducción». Este nombre deriva de la que fue, durante la mayor parte de la historia de la ciencia, la teoría dominante acerca del funcionamiento de ésta. Dicha teoría postulaba que, a falta de demostraciones matemáticas, existe un modo menor, pero todavía suficiente, de justificar las observaciones, denominado *inducción*. La inducción se contraponía, por un lado, con la justificación supuestamente perfecta proporcionada por la *deducción*, y, por otro, con las formas, supuestamente más débiles, desde el punto de vista filosófico, del razonamiento intuitivo, que carecen de evidencias obtenidas de la observación que las respalden. En la teoría inductivista del conocimiento científico, las observaciones tienen dos papeles: primero, en el descubrimiento de teorías científicas, y segundo, en su justificación. Se supone que se descubre una teoría por la «extrapolación» o la «generalización» de los resultados de las observaciones. Así pues, si gran número de observaciones se adaptan a la teoría, y no hay ninguna que discrepe de ella, se la considera justificada, creíble y fiable. Ilustra este esquema la figura 3.1.

El análisis inductivista de lo que he dicho acerca de las sombras sería, pues, el siguiente: «Observamos una serie de sombras y detectamos fenómenos de interferencia (estadio 1). Los resultados se ajustan a los que cabría esperar si existiesen universos paralelos que se afectasen mutuamente en cierta medida. Pero al principio nadie se da cuenta de ello. Con el tiempo (estadio 2), alguien hace la generalización de que las interferencias *siempre* serán observadas si se dan las circunstancias adecuadas, y, por consiguiente, induce la teoría de que los responsables son los universos paralelos. Cada vez que se observan nuevas interferencias (estadio 3), nos convencemos un poco más de la validez de esa teoría. Tras una secuencia suficientemente larga de tales observaciones, y a condición de que ninguna de ellas contradiga la teoría, concluimos (estadio 4) que ésta es cierta. Si bien jamás podremos estar completamente seguros, a efectos prácticos nos damos por convencidos.»

No resulta fácil decidir por dónde empezar la crítica del planteamiento inductivista de la ciencia, dada su profunda falsedad en tan diversos aspectos. Quizás su peor defecto, en mi opinión, sea el evidente *non sequitur* de que una predicción generalizada equivale a una nueva teoría. Como toda teoría científica de cierta profundidad, la de los universos paralelos, simplemente, no tiene la forma de una generalización a partir de la observación. ¿Acaso observamos primero un universo, luego otro, y más tarde un tercero, y de ello inducimos que su número es infinito? ¿Era la generalización de que los planetas «vagarán» por el espacio siguiendo una órbita, y no otra, equivalente a la teoría de que son mundos que giran alrededor del Sol y la Tierra es uno de ellos? Resulta igualmente erróneo pretender que repetir nuestras observaciones sea el modo en que nos convencemos de las teorías científicas. Como he dicho, las teorías son explicaciones, no meras predicciones. Si uno no acepta una explicación propuesta para una serie de observaciones, repetir éstas una y otra vez difícilmente será la solución. Y todavía más difícil será que repetir las observaciones nos ayude a elaborar una explicación satisfactoria cuando no se nos ocurre pensar en ninguna.

Es más, ni siquiera las meras predicciones pueden ser justificadas nunca por la evidencia obtenida mediante la observación, como demostró Bertrand Russell con su historia del pollo. (En evitación de cualquier posible malentendido, permítaseme señalar que se trata de un metafórico pollo antropomorfo que representa a un ser humano que trata de entender las regularidades del universo.) El pollo observa que el granjero va cada día a darle de comer, y predice que lo seguirá haciendo así día tras día. Los inductivistas dirían que el pollo ha «extrapolado» sus observaciones en una teoría, y que cada comida la justifica un poco más. Un buen día, sin embargo, aparece el granjero y, en vez de darle de comer, le retuerce el pescuezo al pollo. El desengaño experimentado por el pollo de Russell lo han sentido también billones de otros pollos. ¡Ello justifica inductivamente la conclusión de que la inducción no puede justificar ninguna conclusión!

Esta crítica, sin embargo, resulta demasiado tolerante con el inductivismo. Ilustra el hecho de que la observación repetida no puede *justificar* teorías, pero, al hacerlo, deja de lado completamente (o, incluso, acepta) una concepción errónea aún

más fundamental: la de que resulta posible extrapolar inductivamente las observaciones para *formar* nuevas teorías. De hecho, es imposible extrapolar observaciones, a menos de haberlas situado previamente dentro de un marco explicativo. A fin de «inducir» su falsa predicción, por ejemplo, el pollo de Russell debe tener previamente una falsa explicación para el i comportamiento del granjero. Quizás suponía que éste tenía sentimientos humanitarios hacia los pollos. Si se le hubiese ocurrido otra explicación—por ejemplo, que el granjero sólo engordaba a sus pollos para comérselos—, habría «extrapolado» su comportamiento de modo bien distinto. Supongamos que, un buen día, el granjero empieza a dar más comida que antes a sus pollos. El modo en que se extrapole este nuevo conjunto de observaciones para predecir el comportamiento del granjero dependerá enteramente de cómo se explique éste. Según la teoría del granjero humanitario, resulta evidente que su benevolencia ha aumentado, y, por lo tanto, los pollos tienen aún menos motivos que antes para preocuparse. Sin embargo, según la teoría del engorde, el nuevo comportamiento resulta sumamente ominoso: es evidente que el momento del sacrificio se acerca.

El hecho de que una misma evidencia procedente de la observación pueda ser «extrapolada» para arrojar dos predicciones diametralmente opuestas —según la explicación que se adopte—, sin que sea capaz de justificar ninguna de ellas, no constituye una limitación accidental del medio agropecuario. Es cierto para toda evidencia fruto de la observación en cualquier circunstancia. Las observaciones no pueden tener ninguno de los dos papeles que les atribuye el esquema inductivista, ni por lo que respecta a las meras predicciones ni, todavía menos, por lo que se refiere a las genuinas teorías explicativas. Sin duda, el inductivismo se basa en una teoría muy sensata del crecimiento del conocimiento —la de que aprendemos de la experiencia—, e históricamente ha estado asociado con la liberación de la ciencia del dogma y la tiranía. Pero si deseamos entender la verdadera naturaleza del conocimiento, así como su lugar en la estructura de la realidad, debemos afrontar el hecho de que el inductivismo es falso de raíz. Ningún razonamiento científico —y, lo que es más, ningún razonamiento, de la clase que sea, que haya resultado cierto—, ha encajado nunca en la descripción inductivista.

¿Cuál es, pues, el patrón del razonamiento y el descubrimiento científicos? Hemos visto ya que el inductivismo, y

las demás teorías del conocimiento centradas en la predicción, se basan en una idea errónea. Lo que necesitamos, entonces, es una teoría del conocimiento centrada en la explicación: una teoría que permita comprender cómo se producen y se justifican las explicaciones; una teoría que permita comprender cómo, por qué y cuándo debemos aceptar que nuestras percepciones cambien nuestra concepción del mundo. Una vez la tengamos, no necesitaremos ya separar la teoría de las predicciones, puesto que, si se tiene la explicación de un fenómeno observable, hacer predicciones acerca de él no es ningún misterio. Y, una vez justificada una explicación, toda predicción derivada de ella lo está también de manera automática.

Por fortuna, la teoría del conocimiento científico que predomina en la actualidad—debida, en gran medida, en su forma moderna, al filósofo Karl Popper, y que constituye una de mis «cuatro vías» principales de explicación de la estructura de la realidad—puede ser considerada una teoría de las explicaciones en el sentido que acabo de exponer. Considera a la ciencia como un proceso de *resolución de problemas*. El inductivismo considera que el conjunto de nuestras observaciones pasadas es una especie de entramado teórico que la ciencia ha de ir rellenando mediante la interpolación y la extrapolación. La resolución de problemas empieza con una teoría inadecuada, pero no con una «teoría» nocional consistente en observaciones pasadas. Se inicia a partir de las mejores teorías disponibles. Si alguna de dichas teorías nos parece inadecuada, y deseamos sustituirla, entonces nos encontramos ante un *problema*. Así pues, y al contrario de lo que propone el esquema inductivista de la figura 3.1, el descubrimiento científico no necesita iniciarse con la evidencia fruto de la observación. Pero siempre empieza con un problema. «Problema» no significa por fuerza una emergencia práctica ni un motivo de ansiedad, sino, simplemente, un conjunto de ideas que parece inadecuado y merece que se intente mejorarlo. La explicación disponible puede parecer demasiado fácil o laboriosa, o innecesariamente estrecha, o ambiciosa de un modo irreal. Quizás vislumbramos alguna posible unificación con otras ideas. O tal vez dos explicaciones, satisfactorias en sus respectivos campos, nos parecen inconciliables. O *puede* que se hayan efectuado interesantes observaciones —como el deambular de los planetas— que las teorías de que disponemos no predijeron y no pueden explicar.

Este último tipo de problema se parece al estadio 1 del esquema inductivista, aunque sólo superficialmente. Y es que una observación inesperada nunca inicia un descubrimiento científico, a menos que la teoría preexistente contenga las semillas del problema. Las nubes, por ejemplo, deambulan más que los planetas. Su impredecible deambular era, lógicamente, familiar antes del descubrimiento de éstos. Por otra parte, la predicción del tiempo siempre ha sido de utilidad para agricultores, marineros y soldados, de modo que siempre debe de haber existido un buen incentivo para teorizar sobre el movimiento de las nubes. Sin embargo, no fue la meteorología la que abrió el camino de la ciencia moderna, sino la astronomía. Sin duda, la evidencia proporcionada por la observación era mucho más abundante en el caso de la meteorología que en el de la astronomía, pero nadie le prestaba demasiada atención ni desarrollaba a partir de ella teorías sobre frentes fríos y anticiclones. La historia de la ciencia no se llenó de disputas, dogmas, herejías, especulaciones y complejas teorías acerca de la naturaleza de las nubes y su movimiento. ¿Por qué? Pues porque, de acuerdo con la estructura explicativa establecida para el tiempo atmosférico, resultaba perfectamente comprensible que el movimiento de las nubes no se podía predecir. El sentido común sugiere que las nubes se mueven a causa del viento. Cuando no lo hacen de acuerdo con el que sopla en superficie, parece razonable deducir que el viento puede soplar en diferentes direcciones a distintas altitudes, por lo que su curso resulta impredecible, y entonces es fácil llegar a la conclusión de que no hay nada más que explicar. Algunas personas adoptaron, sin duda, posturas similares en relación a los planetas y dieron por sentado que no eran más que objetos luminosos en la esfera celeste, impulsados por los vientos de las grandes altitudes, o quizás movidos por ángeles, y que, por lo tanto, no había nada más que explicar. Otros, sin embargo, no se dieron por satisfechos e intuyeron que tras el deambular de los planetas se escondían explicaciones más profundas. Así que fueron en busca de estas explicaciones y las hallaron. En algunos momentos de la historia de la astronomía la cantidad de evidencias inexplicadas fruto de la observación ha sido enorme, en otros, escasa, y en otros, nula. Pero, en todos los casos, si la gente hubiese decidido teorizar en función del número de observaciones acumuladas de un fenómeno determinado, habría optado, sin duda, por las nubes y no por los planetas. Sin embargo, escogió

estos últimos, por muy diversas razones. Algunas de ellas estaban relacionadas con ideas preconcebidas acerca de cómo debía ser la cosmología, o con argumentos expuestos por antiguos filósofos, o con la numerología mística. Otras se basaban en la física, las matemáticas o la geometría de la época. Algunas han resultado estar llenas de mérito objetivo, mientas que otras no.

Pero todas ellas fueron consecuencia de un mismo hecho: en algún momento, a alguien le pareció que las explicaciones de que se disponía podían y debían ser mejoradas.

Se resuelve un problema encontrando nuevas teorías o mejorando las existentes, de modo que contengan explicaciones que no presenten las deficiencias, pero conserven los méritos, de las explicaciones precedentes (figura 3.2). Así, cuando se presenta un problema (estadio 1), el paso siguiente siempre incluye la *conjetura*: la proposición de nuevas teorías o la modificación o reinterpretación de las existentes, con la esperanza de resolver el problema (estadio 2). Las conjeturas son entonces *criticadas*, lo que, si la crítica es racional, conlleva su examen y comparación para ver si realmente ofrecen las mejores explicaciones, según los criterios inherentes al problema (estadio 3). Cuando una teoría conjeturada no sobrevive a la crítica —es decir, cuando parece ofrecer peores explicaciones que otras teorías—, es desechada. Cuando desechamos alguna de las teorías existentes en favor de una nueva (estadio 4), consideramos provisionalmente que nuestro proceso de resolución de problemas ha avanzado. Digo «provisionalmente» porque es probable que la subsiguiente resolución de problemas conlleve la alteración o incluso la sustitución de esas nuevas teorías aparentemente satisfactorias y, en algunas ocasiones, la resurrección de algunas de las que nos parecieron insatisfactorias. Así pues, la solución, por buena que sea, nunca es el fin de la historia, sino el punto de partida para el próximo proceso de resolución de problemas (estadio 5). Esto ilustra otra de las falsedades inherentes al inductivismo. El objetivo de la ciencia no es encontrar una teoría que sea definitiva —o parezca que lo puede ser—, sino encontrar en cada momento la mejor teoría de que podemos disponer y que, a ser posible, mejore todas las teorías existentes. Todo enunciado científico intenta convencernos de que la explicación ofrecida es la mejor de que se dispone. No nos dice, ni nos puede decir, nada acerca de lo que sucederá con esa explicación cuando, en el futuro, se vea sometida a nuevas críticas y comparada con explicaciones que aún

hay que descubrir. Una buena explicación puede hacer buenas predicciones sobre el futuro, pero lo que ninguna de ellas puede ni siquiera sugerir, es el contenido o la calidad de sus rivales futuras.

Lo que he descrito hasta aquí es aplicable a toda resolución de problemas, cualesquiera que sean las materias o las técnicas de crítica racional involucradas. La resolución de problemas *científicos* incluye en todos los casos un método particular de crítica racional, denominado *prueba experimental*. Cuando dos o más teorías rivales hacen predicciones conflictivas sobre el resultado de un experimento, éste es realizado y la teoría o teorías que hicieron falsas predicciones son desechadas. La elaboración de las conjeturas científicas se concentra en el hallazgo de explicaciones que permitan hacer predicciones comprobables experimentalmente. De un modo ideal, andamos siempre a la búsqueda de *pruebas experimentales cruciales*, experimentos cuyos resultados, cualesquiera que sean, harán desechar una o más de las teorías contrastadas. Este proceso se ilustra en la figura 3.3. Tanto si el problema inicial contenía observaciones como si no (estadio 1), y tanto si las teorías contrastadas estaban diseñadas para ser comprobadas experimentalmente como si no (estadio 2), es en la fase crítica del descubrimiento científico (estadio 3) donde las pruebas experimentales tienen un papel característico y decisivo. Dicho papel consiste en demostrar que algunas de las teorías son insatisfactorias poniendo de manifiesto que sus explicaciones conducen a falsas predicciones. Debo mencionar aquí una notable asimetría en la filosofía y la metodología de la ciencia: mientras que una predicción incorrecta convierte automáticamente en insatisfactoria a la explicación en que se basa, una predicción correcta no nos dice nada en absoluto sobre la explicación que la ha originado. Es relativamente frecuente que explicaciones con muy poco fundamento científico hagan predicciones correctas, cosa que deberían tener en cuenta —aunque no lo hacen— los entusiastas de los OVNIS, los teóricos de las conspiraciones y los pseudocientíficos de toda índole.

Si una teoría relativa a hechos observables no es verificable —es decir, si ninguna observación factible está en condiciones de desecharla—, no puede explicar por sí misma por qué esos hechos suceden del modo en que son observados y no de cualquier otro. La teoría de los «ángeles» para justificar el movimiento de los planetas, por ejemplo, no se puede verificar porque, haciendo caso omiso de cómo se muevan las planetas, su movimiento siempre podría ser atribuido a los ángeles; por consiguiente, dicha teoría no puede explicar los movimientos concretos que vemos, a menos que se complemente con una teoría independiente sobre el movimiento de los ángeles. Por esta razón existe una regla metodológica en ciencia que dice que, una vez que una teoría verificable de manera experimental ha superado las pruebas apropiadas, cualquier teoría *menos* verificable acerca de los mismos fenómenos queda automáticamente desechada, puesto que sus explicaciones serán, sin duda, inferiores. Se dice a menudo que esta regla traza la frontera entre la ciencia y otras clases de adquisición de conocimientos. Pero si admitimos que la ciencia trata de explicaciones, veremos que esta regla es, en realidad, un caso especial de algo que se aplica de modo natural a todos los casos de resolución de problemas: *las teorías capaces de ofrecer explicaciones más detalladas son preferidas de modo automático*. Y ello por dos razones: la primera es que una teoría que «se juega el cuello», al ser más específica sobre más fenómenos, se abre a sí misma y a sus rivales a más formas de crítica y, en consecuencia, tiene más probabilidades de hacer avanzar el proceso de resolución de problemas; la segunda es, simplemente, que, si dicha teoría sobrevive a la crítica, dejará menos cosas por explicar, que es de lo que, en realidad, se trata.

He señalado ya que, incluso en ciencia, la mayor parte de la crítica no consiste en pruebas experimentales. Ello es consecuencia de que la crítica científica no se centra en las predicciones de la teoría, sino que va directamente a las explicaciones en que se basan aquéllas. Comprobar las predicciones constituye sólo un modo indirecto (si bien muy importante, cuando resulta posible) de verificar las explicaciones.

En el capítulo 1 ofrecí el ejemplo de la «cura de hierba» (la teoría de que comer un kilo de hierba cura el resfriado común). Esta teoría, y muchísimas más de la misma índole, son fácilmente comprobables, pero podemos criticarlas y desecharlas sin necesidad de realizar ningún experimento, basándonos en el hecho de que no explican nada que represente una novedad en comparación con las teorías existentes y, en cambio, hacen enunciados nuevos que no explican.

Los estadios de un descubrimiento científico que muestra la figura 3.3 rara vez se completan al primer intento. A menudo se dan retrocesos antes de que cada estadio sea completado o, mejor dicho, *solucionado*, ya que cada uno de ellos puede presentar un problema cuya solución requiera por sí misma los cinco estadios de un proceso subsidiario de resolución de problemas. Esto es aplicable incluso al primer estadio, ya que el problema que inicia el proceso no es inmutable. Si no somos capaces de dar con buenas posibles soluciones, podemos vernos obligados a regresar al estadio 1 para reformular el problema, o incluso para escoger un problema distinto. En realidad, su aparente insolubilidad es sólo una de las múltiples razones por las que a menudo nos parece deseable modificar los problemas que estamos resolviendo. Algunas variantes del problema pueden resultar más interesantes o más relevantes para otros problemas; algunas están mejor formuladas; otras parecen potencialmente más fructíferas, o más urgentes, o lo que sea. En muchos casos, la discusión acerca de cuál es exactamente el problema y cuáles serían los atributos de una «buena» explicación es objeto de tantas críticas y conjeturas como las propias soluciones.

De modo similar, si al llegar al estadio 3 la crítica es incapaz de seleccionar la mejor de varias teorías rivales, buscamos nuevos métodos críticos. Si eso no da resultado, podemos volver al estadio 2 y tratar de depurar al máximo las soluciones propuestas (y las teorías de que disponemos), a fin de obtener de ellas más explicaciones y predicciones y facilitar así la detección de los errores. También podemos volver al estadio 1 y tratar de hallar mejores criterios para las explicaciones que iremos encontrando. Y así sucesivamente.

No sólo hay constantes retrocesos, sino que al mismo tiempo los múltiples subproblemas presentan retos constantes a los que hay que hacer frente a medida que surgen. Hasta que el descubrimiento ha sido completado no puede presentarse un

argumento razonablemente secuencial, con un esquema parecido al de la figura 3.3. Puede empezar con la exposición de la última y mejor versión del problema y mostrar acto seguido por qué algunas de las teorías desechadas no superaron la crítica, para presentar a continuación la teoría ganadora, explicar por qué superó la crítica, demostrar que podemos pasarnos sin las teorías desechadas y, finalmente, señalar algunos de los nuevos problemas que plantea el descubrimiento.

Mientras un problema se encuentra en proceso de solución, nos enfrentamos a un grande y heterogéneo cúmulo de ideas, teorías y criterios que compiten por sobrevivir, cada uno de ellos con múltiples variantes. Hay un continuo bullir de teorías, pues las que no son alteradas son sustituidas por otras. Así pues, todas las teorías están sujetas a variación y selección, de acuerdo con criterios que también están sujetos a *variación* y *selección*. En conjunto, este proceso se asemeja a la evolución biológica. El problema se asemeja a un nicho ecológico, y la teoría, a un gen o una especie cuya viabilidad dentro de ese nicho se estuviera comprobando. Las variantes de las teorías, al igual que las mutaciones genéticas, surgen continuamente, y las menos aptas se extinguen al triunfar las más capaces de sobrevivir. El «éxito» es la capacidad para seguir adelante a pesar de la presión selectiva —la crítica— aplicada al nicho concreto. Los criterios para la crítica dependerán, en parte, de las características físicas de cada nicho y, en parte, de los atributos de otros genes y especies —es decir, de otras ideas— ya existentes. Tanto la nueva concepción del mundo que puede estar implícita en una teoría que resuelve un problema como las características distintivas de una nueva especie que ocupa un nicho son propiedades *emergentes* del problema o el nicho. En otras palabras, obtener soluciones es un proceso complejo por definición. No existe un modo sencillo de descubrir la verdadera naturaleza de los planetas a partir (por ejemplo) de una crítica de la teoría de la esfera celeste y algunas observaciones adicionales, al igual que no hay una manera simple de describir el ADN de un koala a partir de las propiedades del eucalipto. La evolución, la prueba y el error —especialmente la forma de prueba y error determinada y correcta que denominamos descubrimiento científico—, son los únicos caminos.

Por esta razón, Popper da el nombre de *epistemología evolutiva* a su teoría de que el conocimiento sólo puede crecer mediante la conjetura y la refutación, según del esquema de la

figura 3.3. Se trata de un importante punto de vista unificador, y, como veremos, se dan otras conexiones entre esas dos vías. No quisiera, sin embargo, destacar únicamente las similitudes entre el descubrimiento científico y la evolución biológica, ya que existen también importantes diferencias entre ambas. Una de ellas es que en biología las variaciones (mutaciones) son aleatorias, ciegas y carentes de propósito, mientras que en la resolución de problemas la creación de nuevas conjeturas es en sí misma un proceso complejo y cargado de conocimiento, guiado por los propósitos de las personas involucradas. Quizás una diferencia más importante aún es que no existe el equivalente biológico de *argumento*. Toda conjetura debe ser probada experimentalmente, lo que constituye una de las razones por las que la evolución biológica es, con mucho, más lenta y menos eficiente. Sin embargo, el vínculo entre ambos procesos es mucho más que una mera analogía: son dos de las «cuatro vías principales», íntimamente relacionadas, que propongo como explicación de la estructura de la realidad.

Tanto en la ciencia como en la evolución biológica, el éxito evolutivo depende de la creación de *conocimiento objetivo* y que sea capaz de sobrevivir, lo que en biología denominamos *adaptación*. Es decir, la capacidad de un gen o una teoría para sobrevivir en un nicho no es una función fortuita de su estructura, sino que depende de que la suficiente información fidedigna y útil se halle o no codificada allí, de modo implícito o explícito. Hablaremos de ello en el capítulo 8.

Podemos ahora empezar a comprender lo que justifica las inferencias que extraemos de nuestras observaciones. No lo hacemos nunca sólo a partir de las observaciones, pero éstas pueden convertirse en significativas en el curso de un argumento cuando ponen de manifiesto deficiencias en alguna de las explicaciones que compiten. Escogemos una teoría científica porque diversos argumentos, de los que únicamente unos pocos dependerán de observaciones, nos han convencido (por un tiempo) de que las explicaciones ofrecidas por todas las demás teorías rivales conocidas son menos verdaderas, menos amplias o menos profundas.

Comparemos unos instantes los esquemas de las figuras 3.1 y 3.3, y consideremos cuán distintas son las concepciones del proceso científico que representan. El inductivismo se basa en observaciones y predicciones, mientras que, en realidad, la ciencia se basa en problemas y explicaciones. El inductivismo presupone

que las teorías se extraen o destilan de algún modo de las observaciones o se justifican por ellas, mientras que, de hecho, las teorías empiezan en la mente de alguien como conjeturas injustificadas que, típicamente, *preceden* a las observaciones que harán desechar las teorías rivales. El inductivismo intenta justificar las predicciones suponiendo que demostrarán su validez en el futuro. La resolución de problemas justifica una explicación por ser mejor que otras de las que se dispone en ese momento. El inductivismo es una continua fuente de toda clase de peligrosos errores porque superficialmente parece plausible, a pesar de su intrínseca falsedad.

Cuando conseguimos la resolución de un problema —científico o de cualquier otra clase—, tenemos a nuestra disposición un conjunto de teorías que, si bien no están exentas de problemas, nos parecen preferibles a aquellas con las que habíamos empezado. Los nuevos atributos que tengan las nuevas teorías dependerán, en consecuencia, de las deficiencias que detectemos en las teorías originales, es decir, en lo que constituía el problema. La ciencia se caracteriza por sus problemas tanto como por su método. Los astrólogos que resuelven el problema de cómo hacer horóscopos lo más interesantes posibles sin arriesgarse a que se puedan demostrar sus errores no contribuyen, ciertamente, ni mucho ni poco a crear conocimientos que puedan ser considerados científicos, aunque hayan utilizado métodos genuinamente científicos (como el estudio de mercados) y se den por satisfechos con la solución. En la auténtica ciencia el problema consiste siempre en la comprensión de algún aspecto de la estructura de la realidad mediante el hallazgo de explicaciones tan amplias y profundas, tan verdaderas y específicas, como sea posible.

Cuando creemos que hemos resuelto un problema, adoptamos, como es natural, la nueva serie de teorías y desechamos las anteriores. Por esta razón, la ciencia, contemplada como búsqueda de explicaciones y resolución de problemas, no presenta ningún «problema de la inducción». No hay ningún misterio en lo que nos hace aceptar momentáneamente una explicación cuando ésta es la mejor que se nos ocurre.

TERMINOLOGÍA

Solipsismo. Teoría de que existe una sola mente y lo que parece realidad exterior no es más que un sueño desarrollado en esa mente. Problema de inducción. Si las teorías no pueden ser justificadas lógicamente por la observación, ¿qué las justifica?

Inducción. Proceso, carente de validez, por el que las teorías generales se suponen obtenidas mediante la acumulación de observaciones, o justificadas por éstas.

Problema. Se presenta un problema cuando alguna de nuestras teorías, especialmente por las explicaciones que proporciona, parece inadecuada y necesitada de mejora.

Crítica. La crítica racional compara teorías rivales con la finalidad de descubrir cuál ofrece las mejores explicaciones, de acuerdo con los criterios inherentes al problema.

Ciencia. El propósito de la ciencia es la comprensión de la realidad mediante explicaciones. El *método crítico* característico (pero no único) empleado en ciencia es la prueba experimental.

Prueba experimental. Experimento cuyo resultado permite desechar

una o varias de las teorías rivales.

SUMARIO

En áreas fundamentales de la ciencia, la observación de efectos cada vez más sutiles y pequeños nos conduce a importantes conclusiones sobre la estructura de la realidad. Estas conclusiones, sin embargo, no pueden ser extraídas por pura lógica de las observaciones. ¿Qué las convierte, pues, en determinantes? Esto es el «problema de la inducción». Según el inductivismo, las teorías científicas son descubiertas mediante la

extrapolación de los resultados de las observaciones y justificadas cuando se obtienen observaciones corroborantes. De hecho, el razonamiento inductivo no es válido, pues resulta imposible extrapolar observaciones a menos que dispongamos previamente de un marco explicativo para ellas. Ahora bien, la refutación del inductivismo, así como la verdadera solución del problema de la inducción, depende de que se admita que la ciencia no es un proceso para hacer predicciones a partir de observaciones, sino para hallar explicaciones. Buscamos explicaciones cuando aquellas de las que disponemos hacen surgir un problema. Nos embarcamos entonces en un proceso de resolución de problemas. Las nuevas teorías empiezan como conjeturas injustificadas, que son criticadas y comparadas según los criterios inherentes al problema. Las que no logran superar la crítica, son desechadas. Las supervivientes se convierten en las nuevas teorías dominantes, algunas de las cuales acaban siendo asimismo problemáticas y nos conducen a buscar mejores explicaciones. En conjunto, el proceso se asemeja a la evolución biológica.

Así pues, adquirimos un conocimiento cada vez más profundo de la realidad mediante la solución de problemas y el hallazgo de mejores explicaciones. Sin embargo, una vez todo ha sido hecho y dicho, juzgar los problemas y las explicaciones pasa a depender de la mente humana, que debe su capacidad de razonar a un cerebro poco fiable y cuya fuente de información son unos sentidos no menos poco fiables. ¿Qué autoriza, pues, a la mente humana a sacar conclusiones sobre la realidad externa, objetiva, a partir de su experiencia y su razón puramente subjetivas?

4. CRITERIOS DE REALIDAD

El gran físico Galileo Galilei, quizás el primero en la moderna acepción del término moderno, realizó múltiples descubrimientos no sólo en física propiamente dicha, sino también en metodología de la ciencia. Revivió la antigua idea de expresar en clave matemática las teorías generales sobre la naturaleza y la mejoró desarrollando el método de la prueba experimental, que caracteriza a la ciencia tal y como la conocemos hoy. Denominó, acertadamente, a estas pruebas *cimenti*, es decir «aquilataciones», en italiano antiguo. Fue uno de los primeros en usar telescopios para el estudio de objetos celestes, y recogió y analizó evidencias en pro de la teoría heliocéntrica, según la cual la Tierra se desplaza siguiendo una órbita alrededor del Sol y gira sobre su propio eje. Es bien conocido por su defensa de dicha teoría y por el duro conflicto con la Iglesia al que le condujo esta defensa. En 1633 la Inquisición le juzgó por herejía y le obligó, bajo amenaza de tortura, a leer en voz alta y de rodillas una larga y abyecta retractación en la que declaraba que «abjuraba, condenaba y detestaba» la teoría heliocéntrica. (Según la leyenda, al ponerse en pie murmuró las palabras «*Eppur si muove*...», es decir, «Y, sin embargo, se mueve...».) A pesar de su retractación, fue condenado a arresto domiciliario, en el que permaneció el resto de sus días. Si bien su castigo fue relativamente leve, no por ello dejó de conseguir su propósito, como explica Jacob Bronowski:

El resultado fue el silencio de los científicos católicos en todas partes a partir de entonces [...] El efecto del juicio y el confinamiento fue la radical interrupción de la tradición científica en el Mediterráneo. (*The Ascent of Man*, página 218).

¿Cómo pudo tener tan graves consecuencias una discusión sobre la configuración del sistema solar? ¿Qué movía a quienes participaron en ella a obrar tan apasionadamente? En realidad, la verdadera discusión no estribaba en si el sistema solar tenía una u otra organización, sino en la brillante defensa por Galileo de una nueva y peligrosa manera de pensar acerca de la realidad. No acerca de la existencia de la realidad, puesto que tanto Galileo

como la Iglesia creían en el *realismo*, el punto de vista inspirado por el sentido común que nos dice que existe realmente un universo exterior y que afecta a nuestros sentidos, incluso si estos sentidos ven aumentada su percepción mediante instrumentos como los telescopios. En lo que difería Galileo era en su concepción de la relación entre la realidad física, por un lado, y las ideas, las observaciones y la razón Humanas, por otro. Defendía que el universo podía ser comprendido en términos de leyes universales formuladas matemáticamente, y que un conocimiento sólido de éstas era accesible a los seres humanos a condición de aplicar su método de formulación, matemática ¿sistemática comprobación .experimental. Según sus propias palabras, «el Libro de la Naturaleza está escrito en símbolos matemáticos». Se trataba de una comparación consciente con ese otro Libro en el que era más convencional confiar.

Galileo consideraba que si su método resultaba efectivamente fiable, como creía, las conclusiones que proporcionaría dondequiera que fuese aplicado serían preferibles a las obtenidas por cualquier otro. Insistía, pues, en que el razonamiento científico debía prevalecer no sólo sobre la intuición y el sentido común, sino también sobre las doctrinas religiosas y la revelación. Era esta idea, y no la teoría heliocéntrica en sí, lo que las autoridades eclesiásticas consideraban peligroso. (Y con razón, puesto que si alguna idea puede reclamar el mérito de haber iniciado la revolución científica y la Ilustración, así como de haber sentado los cimientos seculares de la civilización moderna, es, sin duda, ésta.) De ahí que se prohibiera «sostener o defender» la teoría heliocéntrica *como explicación* del aspecto del cielo nocturno. Pero utilizarla, escribir sobre ella, considerarla «una suposición matemática» o propugnarla como método para hacer predicciones estaba permitido. Por esta razón, la obra de Galileo *Dialogo sopra i due massimi sistemi del mondo*, que comparaba la teoría heliocéntrica con la teoría geocéntrica oficial, había conseguido la autorización de los censores de la Iglesia para la impresión *(imprimatur*). Incluso el propio Papa había manifestado previamente su aquiescencia a que Galileo escribiese semejante libro (si bien en el juicio se utilizó un documento engañoso, que pretendía que se había prohibido con anterioridad a Galileo toda discusión sobre el tema).

Es conveniente, desde el punto de vista histórico, señalar que en tiempos de Galileo no era aún indiscutible que la teoría

heliocéntrica permitiese hacer mejores predicciones que la geocéntrica. Las observaciones disponibles no eran muy exactas. Habían sido propuestas modificaciones *ad hoc* para mejorar la exactitud de la teoría geocéntrica, y no resultaba fácil cuantificar los poderes predictivos de ambas teorías rivales. Además, debe tenerse en cuenta que había más de una teoría heliocéntrica. Galileo creía que los planetas se movían en círculos, mientras que, en realidad, sus órbitas son prácticamente elipses, de modo que los datos tampoco encajaban con la *particular* teoría heliocéntrica que Galileo defendía. (Lo cual, incidentalmente, no explica cómo llegó a convencerse de su bondad gracias a la acumulación de observaciones.) Pero no fue esto lo que hizo que la Iglesia tomara posiciones en esta controversia. No le importaba a la Inquisición dónde *parecían* estar los planetas; lo que de veras le interesaba era la realidad. Quería saber dónde estaban realmente, y, al igual que Galileo, quería hacerlo mediante explicaciones. Los instrumentalistas y los positivistas dirían que, puesto que la Iglesia estaba dispuesta a aceptar las predicciones fruto de la observación de Galileo, toda discusión ulterior era ociosa y su casi inaudible «*Eppur si muove*» no tenía sentido. Pero Galileo sabía qué estaba en juego, al igual que los inquisidores. Al negar la fiabilidad del conocimiento científico, éstos pensaban precisamente en su parte explicativa.

La concepción del mundo de la Iglesia era falsa, pero no ilógica. Sin duda, creía en la revelación y la autoridad tradicional como fuentes fiables de conocimiento, pero tenía también una razón adicional e independiente para criticar la fiabilidad del conocimiento obtenido mediante los métodos de Galileo. Hubiera podido aducir, simplemente, que, por muy abundantes que sean las observaciones o argumentos, jamás podrán probar que una explicación de un fenómeno físico es cierta y otra es falsa. Hubiera podido decir que Dios podía producir aquellos efectos observados repetidamente de infinidad de maneras distintas, de modo que pretender estar en posesión de un método de conocimiento basado meramente en las falibles dotes de observación y razonamiento de un hombre no era más que un vanidoso y arrogante deseo de compararse con el Sumo Hacedor.

Hasta cierto punto, la Iglesia no hacía más que pedir modestia y que se reconociera la falibilidad humana. Y si Galileo declaraba que la teoría heliocéntrica estaba más o menos probada —o a punto de serlo—, desde una perspectiva en cierto modo

inductivista, no les faltaba razón a sus detractores. Si Galileo pretendía que sus métodos podían conferir a cualquier teoría una autoridad comparable a la que la Iglesia reclamaba para sus doctrinas, tenían razón al criticarlo por arrogante (o, como seguramente pensaban, blasfemo), si bien, por supuesto, de acuerdo con los propios criterios que aplicaban, los inquisidores eran aún más arrogantes que él.

¿Cómo podemos, pues, defender a Galileo frente a la Inquisición? ¿Cuál podría ser su defensa ante los cargos de ir demasiado lejos cuando aseguraba que las teorías científicas contienen unconocimiento fiable de la realidad? La defensa popperiana de la ciencia como proceso de resolución de problemas y búsqueda de explicaciones no es por sí misma suficiente. También la Iglesia se interesaba primordialmente por las explicaciones y no por las predicciones, y se mostraba bien dispuesta a permitir que Galileo resolviera problemas utilizando la teoría que quisiese. Lo que no estaba dispuesta a aceptar era que las soluciones de Galileo (lo que denominaba meras «hipótesis matemáticas») tuviesen relación alguna con la realidad extema. Después de todo, la resolución de problemas es un proceso que tiene lugar enteramente en el interior de las mentes humanas. Galileo podía considerar el mundo como un libro en el que las leyes de la naturaleza están escritas en símbolos matemáticos, pero esto, estrictamente hablando, sólo era una metáfora; no hay explicaciones que giren por el espacio en órbita con los planetas. La realidad es que todos.. nuestros problemas, junto con sus soluciones, se encuentran dentro de nosotros, han sido creados por nosotros. Cuando resolvemos problemas científicos, mediante la argumentación elaboramos teorías cuyas explicaciones nos parecen las mejores. De modo que, sin negar en modo alguno que sea justo, adecuado y útil resolver problemas, la Inquisición y los escépticos modernos pueden legítimamente preguntarse qué tiene que ver la resolución de problemas con la realidad. Podemos hallar las «mejores explicaciones» que nos satisfagan desde un punto de vista psicológico. Podemos descubrir que resultan útiles para hacer predicciones. Nos pueden parecer, sin duda, esenciales en cada campo de la creatividad tecnológica. Todo ello justificará que sigamos buscándolas y que las utilicemos para estos fines. Pero ¿por qué hemos de sentirnos obligados a aceptarlas como hechos? Lo cierto es que la proposición que la Inquisición obligó a admitir a Galileo era la siguiente: que la Tierra está, en realidad, en reposo,

y el Sol y los planetas giran a su alrededor; pero los caminos por los que viajan estos cuerpos astronómicos están dispuestos de un modo tan complejo que, vistos desde la perspectiva de la Tierra, resultan coherentes con la concepción de que el Sol está en reposo y la Tierra y los planetas en movimiento. Permítaseme denominarla «teoría de la Inquisición» del sistema solar. Si esta teoría fuese cierta, seguiríamos pudiendo esperar que la teoría heliocéntrica hiciese predicciones ajustadas de los resultados de cualquier observación astronómica realizada desde la Tierra, a pesar de ser factualmente falsa. Así pues, toda observación que pareciera apoyar la teoría heliocéntrica apoyaría también la teoría de la Inquisición.

Podríamos ampliar la teoría de la Inquisición para dar cuenta de observaciones más detalladas que apoyan la teoría heliocéntrica, como las de las fases de Venus y las de los pequeños movimientos adicionales denominados «movimientos propios» de algunas estrellas en relación con la bóveda celeste. Para hacerlo, deberíamos presuponer maniobras aún más complejas en el espacio, gobernadas por leyes físicas muy diferentes de las que se dan en una Tierra que se supone estática. La diferencia sería la estrictamente necesaria para que fueran compatibles con el hecho, demostrado por la observación, de que la Tierra se mueve, y dichas leyes, además, deberían ser válidas tanto en el espacio como aquí abajo. Son posibles muchas teorías así. En efecto, si hacer las predicciones adecuadas fuese nuestra única preocupación, podríamos inventar teorías que dijeran que en el espacio sucede cualquier cosa que nos interese. Por ejemplo, las meras observaciones jamás contribuirían a desechar una teoría que sostuviese que la Tierra está encerrada en un inmenso planetario que nos muestra la simulación de un sistema solar heliocéntrico, y que fuera de ese planetario se encuentra lo primero que se nos ocurra, o nada en absoluto. Por supuesto, para dar cuenta de las observaciones que hacemos cada día, ese planetario debería también redirigir nuestras emisiones de radar y láser, capturar nuestras sondas espaciales y, evidentemente, nuestros astronautas, enviarnos falsos mensajes de dichos astronautas y devolvérnoslos con las adecuadas muestras de suelo lunar, alteraciones de la memoria, etcétera. Tal vez parezca una teoría absurda, pero lo fundamental es que no puede ser desechada por la experimentación. Tampoco es válido desechar ninguna teoría aduciendo únicamente que es «absurda»: en

tiempos de Galileo, no sólo la Inquisición, sino casi todo el mundo, consideraba el colmo del absurdo afirmar que la Tierra se mueve. Después de todo, no notamos su movimiento, ¿verdad? Cuando se mueve, durante en un terremoto, por ejemplo, lo notamos inequívocamente. Se dice que Galileo retrasó unos años su defensa pública de la teoría heliocéntrica no por miedo a la Inquisición, sino al ridículo.

La teoría de la Inquisición nos parece, sin duda, absurdamente artificial. ¿Por qué deberíamos aceptar una explicación tan complicada y *ad hoc* para justificar el aspecto del cielo cuando la sencilla cosmología heliocéntrica hace lo mismo y sin tantas complicaciones? Podríamos aducir el principio de la navaja de Occam: «No multiplicar las entidades más allá de lo necesario.» O, como prefiero decir: «No complicar las explicaciones más allá de lo necesario», ya que, si lo hacemos, las propias complicaciones innecesarias quedarán sin explicar. Sin embargo, que una explicación sea considerada o no «artificial» o «innecesariamente complicada» depende del conjunto de ideas y explicaciones que constituyan la concepción del mundo en cada persona en un momento dado. La Inquisición habría podido argumentar que la idea de una Tierra en movimiento resultaba innecesariamente complicada. Contradecía el sentido común, contradecía las Sagradas Escrituras y (hubieran podido añadir los inquisidores) existía ya otra explicación perfectamente válida y que no necesitaba recurrir a semejante idea.

Pero ¿existía realmente? ¿Proporcionaba la teoría de la Inquisición explicaciones alternativas sin tener que recurrir a la «complicación», tan contraria al sentido común, del sistema heliocéntrico? Echemos una atenta mirada al modo en que la teoría de la Inquisición explica las cosas. Justifica la aparente inmovilidad de la Tierra afirmando que *está* inmóvil. Hasta aquí, muy bien. A primera vista, esta explicación es mejor que la de Galileo, quien tuvo que emplearse a fondo y rebatir algunas nociones de sentido común sobre fuerza e inercia para explicar por qué no notamos el movimiento de la Tierra. Pero ¿cómo se enfrenta la teoría de la Inquisición a la tarea, y más difícil, de explicar los movimientos de los planetas?

La teoría heliocéntrica los explica diciendo que vemos que los planetas siguen complicadas trayectorias por el espacio porque, aunque en realidad describen círculos (o elipses), mientras lo hacen la Tierra también se mueve. La explicación de la Inquisición

consiste en que los planetas parecen moverse trazando complicados circuitos por el espacio porque eso es precisamente lo que hacen; *pero* (y en ello, según la teoría de la Inquisición, estriba la esencia de la explicación) este complejo movimiento es gobernado por un principio sencillo, pero fundamental: los planetas se mueven de tal modo que, vistos desde la Tierra, parece que tanto ellos como ésta se desplacen siguiendo órbitas sencillas alrededor del Sol.

Para comprender el movimiento de los planetas en términos de la teoría de la Inquisición es imprescindible entender este principio fundamental, puesto que las limitaciones que impone constituyen la base de toda explicación detallada que se desee formular de acuerdo con ella. Por ejemplo, si nos preguntasen por qué se da una determi nada conjunción de planetas en una fecha concreta, o por qué un determinado planeta describe cierta trayectoria particular a través del cielo, la respuesta sería siempre: «Porque así es como se vería si la teoría heliocéntrica fuese cierta.» Nos encontramos, pues, ante una cosmología —la de la Inquisición— que sólo puede ser comprendida en términos de otra distinta —la heliocéntrica—, a la que contradice, pero que imita fielmente.

Si los inquisidores hubiesen tratado en serio de comprender el mundo en los términos de la teoría que intentaron imponer a Galileo, también se habrían dado cuenta de su fatal debilidad, es decir, que es incapaz de solucionar el problema que pretende resolver. *No* explica las evoluciones de los planetas «sin necesidad de introducir la complicación del sistema heliocéntrico», sino que, al contrario, lo incorpora inevitablemente como parte de su propio principio fundamental para explicar el movimiento de los planetas. No se puede comprender el mundo mediante la teoría de la Inquisición a menos que se entienda primero la teoría heliocéntrica.

Tenemos razón, pues, al considerar a la teoría de la Inquisición como una retorcida elaboración de la teoría heliocéntrica, en vez de lo contrario. No hemos llegado a esta conclusión juzgando la teoría de la Inquisición de acuerdo con la cosmología moderna, lo que habría sido una falacia, sino insistiendo en tomarla en serio, en sus propios términos, como explicación del mundo. He mencionado anteriormente la teoría de la cura de hierba, que podemos desechar sin necesidad de someterla a pruebas experimentales, puesto que no contiene

explicación alguna. La de la Inquisición es otra teoría que podemos desechar del mismo modo; sin necesidad de comprobarla experimentalmente puesto que contiene una mala explicación, una explicación que, en sus propios términos, es peor que su rival.

Como he dicho, los inquisidores eran realistas. Sin embargo, su teoría tiene punto en común con el solipsismo: ambas teorías trazan una frontera arbitraria, más allá de la cual —según afirman— no puede llegar la razón humana. O, por lo menos, más allá de esa frontera la resolución de problemas no sirve para comprender la realidad exterior. Para los solipsistas, esa frontera se limita a sus propios cerebros, o quizás únicamente a sus mentes abstractas o sus almas incorpóreas. Para la Inquisición, incluía toda la Tierra. Algunos creacionistas actuales creen en una frontera similar, no en el espacio, sino en el tiempo, ya que están convencidos de que sólo hace seis mil años que el universo fue creado en su totalidad, lo que incluye incluso las evidencias irrefutables de acontecimientos anteriores. El *conductismo* sostiene que carece de sentido intentar explicar el comportamiento humano en términos de procesos mentales internos. Para los conductistas la única psicología legítima es el estudio de las respuestas observables de las personas al recibir estímulos externos. Delinean así la misma frontera que los solipsistas al separar la mente humana de la realidad exterior, pero mientras que éstos dicen que carece de importancia pensar en cualquier cosa que quede fuera de dicha frontera, aquéllos dicen que lo que no tiene importancia es pensar en cualquier cosa que se halle dentro de ella.

Existe gran variedad de teorías similares, pero, a efectos prácticos, podemos considerarlas variantes del solipsismo. Difieren entre sí en dónde sitúan el límite de la realidad (o de aquella parte de la realidad comprensible mediante la resolución de problemas). Difieren también en si, y cómo, buscan el conocimiento fuera de dicha frontera. Pero todas se asemejan en que consideran que la racionalidad científica y otras maneras de resolver problemas son inaplicables más allá del límite establecido, y hacerlo no para de ser un juego. Pueden llegar a conceder que se trate de un juego satisfactorio y útil, pero sigue sin ser más que un juego, del que no puede sacarse ninguna conclusión válida sobre la realidad exterior.

También se asemejan en que comparten una objeción fundamental a la resolución de problemas como medio de creación de conocimiento, consistente en argumentar que ésta no extrae

sus conclusiones de ninguna fuente última de justificación. Dentro de las fronteras que establece cada una de ellas, los seguidores de todas esas teorías confían en la metodología de la resolución de problemas, pues no ignoran que buscar la mejor explicación posible es también el modo de encontrar la mejor teoría posible. Pero en cuanto a la veracidad de lo que queda fuera de dichas fronteras, todos ellos miran hacia otra parte en busca de una fuente última de justificación. Para las personas religiosas, la revelación divina puede tener este papel. Los solipsistas confían únicamente en la experiencia directa de sus propios pensamientos, tal como los expresó René Descartes con su clásico argumento *Cogito, ergo sum* («Pienso, luego existo»),

A pesar de su deseo de apoyar su filosofía en esta base supuestamente sólida, Descartes se permitía muchas otras hipótesis y suposiciones, y no era, ni mucho menos, solipsista. En realidad, debe de haber habido bien pocos solipsistas genuinos en la historia, si es que ha habido alguno. El solipsismo sólo es defendido, habitualmente, como medio para atacar al razonamiento científico, o como base de partida para alguna de sus múltiples variantes. Por la misma razón, un buen modo de defender a la ciencia contra gran variedad de críticas, así como de comprender la verdadera relación entre razón y realidad, es considerar el argumento contra el solipsismo.

Existe un conocido chiste filosófico sobre un profesor que en clase hace una defensa del solipsismo. Ésta resulta tan convincente, que, en cuanto termina, varios estudiantes entusiastas corren a felicitarlo. «Fantástico, estoy de acuerdo con todo lo que ha dicho», asegura uno. «Yo también», afirma otro. «Me satisface mucho oírlo», responde el profesor. «Pocas veces se tiene la oportunidad de encontrarse con otros solipsistas.»

El chiste lleva implícito un buen argumento genuino contra el solipsismo. Podríamos plantearlo como sigue: ¿en qué consistía, exactamente, la teoría con la que los estudiantes manifestaban estar de acuerdo? ¿Era la teoría del profesor de que ellos no existían porque sólo existía él? De ser así, hubiesen debido encontrar primero un medio de soslayar el *Cogito, ergo sum* cartesiano. De haberlo conseguido, ya no serían solipsistas, puesto que la tesis central del solipsismo es que el solipsista existe. ¿Acaso había sido persuadido cada estudiante de una teoría que *contradecía* la del profesor, la teoría de que aquel estudiante concreto existía, pero no los demás ni el profesor? Ello los habría

convertido a todos en solipsistas, sin duda, pero entonces ninguno de ellos estaría de acuerdo con la teoría que defendía el profesor. Así pues, ninguna de estas dos posibilidades significa que los estudiantes hayan quedado convencidos por la defensa del solipsismo hecha por el profesor. Si adoptan la opinión de éste, no serán solipsistas, y si se convierten en solipsistas, por fuerza deberán tener la convicción de que el profesor está equivocado.

Este razonamiento trata de poner en evidencia que no se puede defender el solipsismo, puesto que quien lo hace lo contradice implícitamente. Pero nuestro solipsista profesor podría tratar de soslayar este argumento diciendo algo así como: «Puedo defender coherentemente el solipsismo, y así lo hago. No contra otros individuos, ya que no existe ningún individuo más, sino contra los razonamientos contrarios. Éstos llegan hasta mí a través de individuos ilusorios que se comportan igual que si fuesen seres pensantes con ideas opuestas a las mías. Mi clase, y los razonamientos que defendí en ella, no se proponían persuadir a esos individuos ilusorios, sino que iban destinados a mí, para ayudarme a aclarar mis ideas.»

Sin embargo, si hay fuentes de ideas que se comportan *como si* fuesen independientes de mí, es que *son* necesariamente independientes de mí. Si defino mi «yo» como el ente consciente que alberga los pensamientos y sentimientos de los que soy cabal, los individuos ilusorios con los que parezco interactuar son, por definición, algo distinto que ese «yo» tan estrechamente definido, así que debo admitir la existencia de algo más que «yo». La única opinión alternativa que tendría, si fuese un solipsista convencido, sería considerar a los individuos ilusorios como creaciones de mi mente inconsciente y, por lo tanto, parte de mi «yo» en un sentido más amplio. Pero entonces debería admitir que mi «yo» estaba dotado de una estructura muy rica y, en su mayor parte, independiente de mi ser consciente. Dentro de esta estructura, habría entes—individuos ilusorios— que, a pesar de ser meros constituyentes de la mente de un supuesto solipsista, se comportarían como si fuesen *anti*solipsistas convencidos. No podría, pues, considerarme completamente solipsista, ya que sólo mi «yo» estrechamente definido sostendría ese punto de vista. Muchas—prácticamente, la mayoría—de las opiniones contenidas en el conjunto de mi mente se opondrían al solipsismo. Podría estudiar la región «exterior» de mi «yo» y descubrir que parece obedecer a ciertas leyes, las mismas que los libros de texto

ilusorios dicen ser de aplicación a lo que llaman el universo físico. Podría descubrir que esa región exterior es mucho más rica que la interior. Además de contener más ideas, es también más compleja y más variada, y está dotada de infinitamente más variables mensurables.

Esa región exterior, además, es susceptible de ser estudiada científicamente utilizando los métodos de Galileo. Puesto que me he visto obligado a definir esa región como parte de mi yo, al solipsismo no le queda ya ningún argumento para oponerse a ese estudio, que ahora habría que calificar de mera forma de introspección. El solipsismo permite, e incluso asume, que el conocimiento de uno mismo puede ser obtenido mediante la introspección. No puede declarar irreales a los entes y procesos objeto de estudio, puesto que la realidad del yo es su postulado básico.

Como podemos ver, si nos tomamos el solipsismo en serio —si asumimos que es cierto y que toda explicación válida debe adaptarse escrupulosamente a él—, se destruye a sí mismo. ¿En qué difiere exactamente el solipsismo, tomado en serio, del realismo, su rival, que se basa en el sentido común? La diferencia se basa tan sólo en una extraña afición del solipsismo a cambiar los nombres de las cosas. Se empecina en referirse a cosas objetivamente distintas (como la realidad externa y mi mente inconsciente, o la introspección y la observación científica), utilizando los mismos nombres. Pero llega el momento en que se ve obligado a reintroducir las distinciones, y lo hace mediante explicaciones que utilizan expresiones como «la parte externa de mi yo». Ninguna de estas explicaciones adicionales sería necesaria de no ser por esa inexplicable insistencia en cambiar los nombres de las cosas. El solipsismo se ve obligado a postular también la existencia de una clase adicional de procesos, unos procesos invisibles e inexplicables, que proporcionan a la mente la ilusión de vivir en una realidad exterior. El solipsista, que cree que nada existe más allá del contenido de una sola mente, la suya, debe creer también que esta mente es un fenómeno de mayor multiplicidad de lo que se supone normalmente. Contiene pensamientos que son como otros individuos, pensamientos que son como los planetas y pensamientos que son como las leyes de la física. Estos pensamientos son reales. Se desarrollan de un modo complejo (o lo simulan) y tienen autonomía suficiente para sorprender, decepcionar, ilustrar o incluso frustrar a esa otra clase

de pensamientos que se denominan a sí mismos «yo». Así pues, el solipsismo explica el mundo como una interacción de pensamientos y no de objetos. Estos pensamientos, sin embargo, son reales e interactúan de acuerdo con las mismas leyes que, según los realistas, rigen la interacción entre los objetos. Así pues, el solipsismo, lejos de constituir una concepción del mundo despejada de alharacas y reducida a lo esencial, es, en realidad, una forma disfrazada de realismo, cargada con un inútil equipaje de asunciones adicionales innecesarias que han sido introducidas sólo para contrarrestar mediante argumentos convincentes las objeciones que se le hacen.

De acuerdo con este razonamiento, podemos prescindir del solipsismo y todas las teorías relacionadas con él, puesto que son indefendibles. Incidentalmente, ya hemos rechazado otra concepción del mundo, el positivismo (es decir, la teoría de que todo enunciado que no describa o prediga observaciones carece de significado), por la misma razón. Como señalé en el capítulo 1, el positivismo enuncia su propia carencia de significado, y no puede, por tanto, ser defendido coherentemente.

Podemos seguir adelante, pues, apoyándonos en el realismo fruto del sentido común y la búsqueda de explicaciones mediante métodos científicos. Pero, a la luz de esta conclusión, ¿qué podemos decir de los argumentos que hacían que el solipsismo y las teorías relacionadas con él parecieran superficialmente plausibles, es decir, los de que su falsedad no podía probarse y no era posible desecharlos mediante la experimentación? ¿Cuál es el status actual de esos argumentos? Si no hemos probado la falsedad del solipsismo, ni lo hemos desechado mediante la experimentación, *¿qué* hemos hecho?

Esta pregunta lleva implícitamente la asunción de que las teorías pueden ser clasificadas según una jerarquía («matemáticas» → «científicas» → «filosóficas») de fiabilidad intrínseca decreciente. Son legión las personas que dan esta jerarquía por sentada, a pesar del hecho de que los juicios acerca de la relativa fiabilidad de las teorías se basan exclusivamente en argumentos filosóficos, que de por sí son bien poco fiables. La propia idea de esa jerarquía es, en realidad, pariente cercana del error reduccionista discutido en el capítulo 1 (la teoría de que las leyes y los fenómenos microscópicos son más fundamentales que las leyes y los fenómenos emergentes). La misma asunción se da en el inductivismo, que afirma que podemos estar absolutamente

seguros de las conclusiones matemáticas, por ser deductivas, y razonablemente seguros de las científicas, por ser «inductivas», pero que debemos dudar siempre de las conclusiones filosóficas, que considera poco más que cuestiones de gusto.

Nada de ello es cierto. Las explicaciones no son justificadas por los medios por los que se llega a ellas, sino por su superior capacidad, en comparación con sus rivales, para solucionar el problema al que se enfrentan. Ésta es la razón de que el argumento de que una teoría es *indefendible* resulte tan concluyente. Una predicción o un enunciado indefendibles pueden, sin embargo, seguir siendo ciertos, pero una explicación indefendible no es una explicación. El rechazo de las «meras» explicaciones basándose en que no están justificadas por explicaciones *definitivas*, nos conduce inevitablemente a búsquedas inútiles de una fuente definitiva de justificación. No existe tal fuente.

Tampoco existe esa supuesta jerarquía de fiabilidad que va de lo matemático a lo filosófico, pasando por lo científico. Algunos argumentos filosóficos, incluyendo el utilizado contra el solipsismo, son mucho más concluyentes que cualquier argumento científico. De hecho, cualquier argumento científico presupone la falsedad, no sólo del solipsismo, sino de otras teorías filosóficas, incluyendo cualesquiera variantes del solipsismo que pudieran contradecir aspectos específicos de ese argumento científico. Demostraré también en el capítulo 10 que incluso los argumentos puramente matemáticos derivan su fiabilidad de las teorías físicas y filosóficas que los apoyan y que, por lo tanto, no pueden, en última instancia, garantizar una absoluta certeza.

Una vez adoptado el realismo, continuamente hemos de decidir si las entidades a las que se refieren las explicaciones rivales son reales o no. Decidir que no lo son —como hicimos en el caso de la teoría de los «ángeles» para explicar el movimiento de los planetas— equivale a rechazar la correspondiente explicación. Así pues, al buscar explicaciones y juzgarlas necesitamos algo más que una simple refutación del solipsismo. Necesitamos desarrollar razones para aceptar o rechazar la existencia de las entidades que puedan aparecer en las teorías contendientes. En otras palabras, necesitamos un *criterio de realidad*. No deberíamos, desde luego, aspirar a encontrar un criterio definitivo o infalible. Nuestro juicio acerca de lo que es real o no, siempre depende de las diversas explicaciones disponibles en cada momento, y a veces cambia a

medida que éstas mejoran. En el siglo XIX, por ejemplo, pocas cosas habrían sido consideradas reales con más confianza que la fuerza de la gravedad. No sólo formaba parte del por entonces sin rival sistema de leyes de Newton, sino que todo el mundo podía sentirla en cualquier momento, incluso con los ojos cerrados (o, al menos, eso creían). Hoy día comprendemos la fuerza de la gravedad por medio de la teoría de Einstein, en lugar de la de Newton, y sabemos que no existe tal fuerza. ¡Y *no* la sentimos! Lo que sentimos es la resistencia que nos impide penetrar en el sólido suelo que hay bajo nuestros pies. Nada tira de nosotros hacia abajo. La única razón de que nos caigamos es que la estructura del tiempo y del espacio en que existimos es curva.

No sólo cambian las explicaciones, sino que nuestros criterios e ideas acerca de lo que debe ser considerado explicaciones también cambian (y mejoran) gradualmente. En consecuencia, la lista de los modos de explicación admisibles permanecerá siempre abierta, así como la de los criterios de realidad aceptables. Pero ¿qué es lo que hace que una explicación—suponiendo que, por las razones que sea, la encontremos satisfactoria— nos mueva a considerar reales a ciertas cosas e ilusorias o imaginarias a otras?

James Boswell relata en su biografía del doctor Johnson que un día estaban discutiendo la teoría solipsista del obispo Berkeley acerca de la inexistencia del mundo material. Boswell comentó que, si bien nadie creía en ella, tampoco era posible refutarla. El doctor Johnson dio una patada a una gran piedra y dijo, mientras su pie rebotaba: «La refuto *así.*» El argumento del doctor Johnson era que la negación de Berkeley de la existencia de la piedra era incompatible con encontrar una explicación al rebote de su pie, que él acababa de sentir. El solipsismo no puede explicar por qué aquel experimento—o cualquier otro—debía tener un determinado desenlace y no otro. Para explicar el efecto que la piedra tuvo sobre él, el doctor Johnson se vio obligado a adoptar una postura acerca de la naturaleza de las piedras. ¿Eran parte de una realidad externa autónoma, o bien quimeras de su imaginación? En este último supuesto, hubiera debido concluir que su «imaginación» constituía de por sí un universo, vasto, complejo y autónomo. Al mismo dilema se habría enfrentado el profesor solipsista, si, en el caso de verse forzado a dar explicaciones, hubiera tenido que tomar una posición en relación con la naturaleza de sus alumnos. Igualmente, la Inquisición se habría visto obligada a tomar posición en relación con la fuente de la regularidad que caracteriza

el movimiento de los planetas, una regularidad sólo explicable haciendo referencia a la teoría heliocéntrica. A todos ellos, el hecho de asumir en serio su propia posición como una explicación del mundo los conduciría directamente al realismo y a la racionalidad de Galileo.

Pero la idea del doctor Johnson es más que una refutación del solipsismo. Ilustra también el criterio de realidad utilizado por la ciencia, es decir, *si algo puede devolver el golpe, existe*. «Devolver el golpe» no significa aquí, necesariamente, que el objeto en cuestión responda al hecho de ser golpeado, de verse afectado desde un punto de vista físico del doctor Johnson. Basta con que cuando «golpeamos» algo, el objeto nos afecte de un modo que requiera una explicación independiente. Galileo, por ejemplo, no tenía manera de afectar a los planetas, pero podía afectar a la luz que le llegaba desde ellos. Su equivalente de la patada a la piedra fue refractar dicha luz a través de las lentes de sus telescopios y de sus ojos. La luz respondió «golpeando» su retina. El modo en que lo hizo le permitió deducir no sólo que la luz era real, sino que los movimientos helicocéntricos de los planetas necesarios para explicar el modo en que la luz llegaba hasta él eran reales también.

Por cierto, el doctor Johnson ni siquiera dio la patada de un modo directo. Una persona es una mente, no un cuerpo. El doctor Johnson que realizó el experimento era una mente, y dicha mente sólo «golpeó» de manera directa unos nervios, los cuales transmitieron unas señales a unos músculos, que impulsaron su pie hacia la piedra. Poco después, el doctor Johnson tuvo la sensación de ser «golpeado a su vez» por la piedra, pero, de nuevo, sólo indirectamente, después de que el impacto hubiera transmitido una sensación de presión a su zapato y éste a su piel, lo que originó nuevos impulsos en sus nervios, etcétera, etcétera. La mente del doctor Johnson, como la de Galileo y la de cualquier otro, «golpeó» nervios y fue «golpeada» por ellos, e infirió la existencia de la realidad y sus propiedades sólo a partir de esas interacciones. *Lo que* el doctor Johnson podía inferir acerca de la realidad dependía de la mejor o peor explicación que fuera capaz de dar de lo que había ocurrido. Si, por ejemplo, le parecía que aquella sensación no era consecuencia de faetores externos, sino sólo de la extensión de su pierna, podía haber concluido que se trataba de una propiedad de su pierna, o únicamente de su mente. Podía padecer una enfermedad que le hiciese sentir una sensación

de rebote cada vez que extendía la pierna de determinada manera. Pero, en realidad, el rebote dependía de lo que hiciera la piedra, como estar en cierto lugar, lo cual estaba relacionado, a su vez, con otros efectos de la piedra, como ser visible o afectar a otras personas que la golpeasen. El doctor Johnson percibió esos efectos como algo autónomo (independientes de él) y bastante complejo. Así pues, la explicación realista de por qué la piedra produce la sensación de rebote incluye una compleja historia sobre algo autónomo. Pero lo mismo haría la explicación solipsista. De hecho, cualquier explicación que tenga en cuenta el fenómeno del rebote del pie debe, necesariamente, ser una «compleja historia sobre algo autónomo». Debe, en efecto, ser la historia de la piedra. Los solipsistas la denominarían una piedra ilusoria, pero, aparte de esto, la historia solipsista y la realista podrían compartir el mismo guión.

Mi exposición sobre las sombras y los universos paralelos del capítulo 2 giraba alrededor del problema de lo que existe o no e, implícitamente, de lo que debería contar o no como evidencia de que algo existe. Para ello utilicé el criterio del doctor Johnson. Consideremos de nuevo el punto X de la pantalla de la figura 2.7 (página 52), que se ilumina cuando sólo dos ranuras están abiertas, pero se oscurece cuando se abren dos más. Dije que la conclusión de que por el segundo par de ranuras pasaba *algo* que impedía que la luz del primer par llegara a X era «inescapable». No es que sea *lógicamente* inescapable, ya que, si no estuviésemos buscando explicaciones, podríamos limitarnos a decir que los fotones que vemos se comportan *como si* algo que pasase por las otras ranuras los desviase, pero que, en realidad, no ocurre nada de eso. De modo semejante, el doctor Johnson hubiese podido afirmar que su pie rebotaba *como si* allí hubiese una piedra, pero que, en realidad, no había nada. La Inquisición mantenía que los planetas se veían *como si* ellos y la Tierra estuviesen evolucionando en órbitas alrededor del Sol, pero que, en realidad, se movían alrededor de la Tierra. Sin embargo, si lo que nos proponemos es explicar el movimiento de los planetas o los fotones, debemos obrar como el doctor Johnson. Debemos adoptar como regla metodológica que el hecho de que algo se comporte como si existiese, es decir, que «devuelva el golpe», ha de ser considerado una evidencia de que existe. Los fotones fantasmas «devuelven el golpe» interfiriendo a los fotones tangibles, luego existen.

¿Podemos concluir del mismo modo, de acuerdo con el criterio del doctor Johnson, que «los planetas se mueven como si fuesen empujados por ángeles, luego éstos existen»? No, pero sólo porque disponemos de una explicación mejor. La teoría de que los ángeles son responsables del movimiento de los planetas no está *del todo* desprovista de mérito. Explica por qué éstos se mueven con independencia de la bóveda celeste, lo que hace, ciertamente, que sea superior al solipsismo. Sin embargo, no explica por qué tendrían que empujar los ángeles a los planetas por unas órbitas y no por otras ni, en particular, por qué tendrían que hacerlo como si su movimiento estuviese determinado por la curvatura del espacio y el tiempo, especificada con todo detalle por las leyes universales de la teoría general de la relatividad. Por esta razón la teoría de los ángeles no puede competir, como explicación, con las teorías de la física moderna.

De modo semejante, postular que unos ángeles pasan por las ranuras adicionales y desvían los fotones visibles sería mejor que nada, pero disponemos de explicaciones más convincentes. Sabemos cómo deberían comportarse esos ángeles: exactamente, como fotones. Así pues, podemos escoger entre una explicación en términos de ángeles que pretenden ser fotones y otra en términos de fotones invisibles. En ausencia de una explicación independiente de por qué tendrían los ángeles que pretender ser fotones, la última explicación resulta superior.

No notamos la presencia de nuestras contrapartidas en otros universos. Tampoco notaba la Inquisición que la Tierra se movía bajo sus pies, y, sin embargo, ¡se mueve! Consideremos ahora cómo sería nuestra existencia si hubiese múltiples copias de cada uno de nosotros que interactuasen tan sólo a través de los efectos imperceptiblemente débiles de la interferencia cuántica. Esto es el equivalente de lo que hizo Galileo cuando analizó qué notaríamos si la Tierra se estuviese moviendo de acuerdo con la teoría heliocéntrica. Descubrió que el movimiento sería imperceptible, si bien «imperceptible» quizá no es la palabra más adecuada. Ni el movimiento de la Tierra ni la presencia de universos paralelos son directamente perceptibles, pero, de hecho, nada lo es (excepto, tal vez, la mera existencia de cada uno de nosotros, si el argumento de Descartes es cierto). Pero ambas cosas son perceptibles, en el sentido de que nos «devuelven el golpe» si las examinamos mediante instrumentos científicos. Podemos ver cómo un péndulo de Foucault oscila sobre un plano que parece desplazarse

gradualmente, lo que manifiesta la rotación de la Tierra bajo él. Podemos detectar fotones que han sido desviados por la interferencia de sus contrapartidas de otros universos. Tan sólo un accidente de la evolución, por así decirlo, hace que los sentidos con los que nacemos no nos permitan notar esos fenómenos directamente.

No es la fuerza con que algo nos «devuelve el golpe» lo que hace que sea convincente la teoría de su existencia. Lo que realmente importa es su papel en las explicaciones que dicha teoría proporciona. He ofrecido ejemplos sacados de la física en los que «golpes» muy pequeños nos conducen a conclusiones muy importantes sobre la realidad porque no disponemos de otra explicación. Lo contrario puede ocurrir también: si no hay una clara vencedora entre las explicaciones rivales, ni siquiera un poderoso «golpe» podrá convencernos de que la supuesta fuente tiene realidad independiente. Un buen día, por ejemplo, vemos que nos atacan terribles monstruos... y luego nos despertamos. Si la explicación de que se originaron en nuestras mentes nos parece adecuada, resultará irracional deducir que existen realmente semejantes monstruos y nos acechan. Si sentimos un súbito dolor en el hombro mientras caminamos por la calle, miramos a nuestro alrededor y no vemos nada que lo explique, podemos considerar si ha sido causado por alguna parte inconsciente de nuestra mente, o por nuestro cuerpo. También podemos considerar *posible* que algún gamberro nos haya disparado con una carabina de aire comprimido, lo cual, sin embargo, no nos permidrá llegar a ninguna conclusión sobre la realidad de dicha persona. Pero si entonces vemos un balín rodando por el suelo, concluiremos que ninguna explicación soluciona mejor el problema que la de la carabina de aire comprimido, y, evidentemente, la adoptaremos. En otras palabras, deduciremos la existencia de alguien a quien no hemos visto, y a quien quizás nunca veremos, sólo porque la intervención de esa persona es la mejor explicación disponible. Está claro que la teoría de la existencia de esta persona no es una consecuencia lógica de la evidencia observada (que, incidentalmente, consistiría en una única observación). Dicha teoría tampoco reviste la forma de una «generalización inductivista», por ejemplo, la de que observaríamos lo mismo si repitiésemos el experimento. La teoría tampoco es comprobable de modo experimental: la experimentación nunca probaría la no intervención de un gamberro oculto. A pesar de todo, el

argumento en favor de la teoría sería abrumadoramente convincente, sin duda, si se tratase de la mejor explicación.

Siempre que he usado el criterio del doctor Johnson para defender la realidad de algo, un atributo en particular se ha destacado: *la complejidad*. Preferimos las explicaciones sencillas a las complicadas, y preferimos las que son capaces de dar cuenta de los detalles y las cuestiones complejas a las que sólo ilustran aspectos simples de los fenómenos. El criterio del doctor Johnson nos dice que consideremos reales estas complejas entidades, ya que, de *no* hacerlo, se complicarían nuestras explicaciones. Debemos, por ejemplo, considerar que los planetas son reales porque, de no hacerlo, nos veríamos forzados a idear complicadas explicaciones basadas en la existencia de un planetario cósmico, o en leyes físicas alteradas, o en ángeles, o en cualquier otra cosa que, de acuerdo con dicha asunción, nos proporcionase la ilusión de que hay planetas en el espacio exterior.

Así pues, la complejidad observada en la estructura o el comportamiento de una entidad es parte de la evidencia de que dicha entidad es real, pero no resulta suficiente. No podemos, por ejemplo, considerar que nuestro reflejo en un espejo es una persona real. Las ilusiones son, sin duda, procesos físicos reales, pero las entidades ilusorias que nos ofrecen no tienen por qué ser consideradas reales, ya que su complejidad deriva de otra fuente. No son *autónomamente* complejas. ¿Por qué aceptamos la teoría del «espejo» para los reflejos, y no la del «planetario» para el sistema solar? Porque, mediante una sencilla explicación sobre la acción de los espejos, podemos entender que nada de lo que observamos en ellos está realmente detrás de ellos. No es necesaria ninguna explicación adicional, puesto que los reflejos, por complejos que sean, nunca serán autónomos: simplemente, toman prestada su complejidad de nuestro lado del espejo. No ocurre lo mismo con los planetas. La teoría de que el planetario cósmico es real y no hay nada más allá de él sólo empeora el problema, ya que, si la aceptamos, en lugar de preguntar cómo funciona el sistema solar, deberemos preguntar primero por el funcionamiento del planetario y luego por el del sistema solar que muestra. No podríamos soslayar esta última pregunta, que es, evidentemente, repetición de la que hacíamos en primer lugar. Estamos ahora en condiciones de reformular el criterio del doctor Johnson como sigue: «Si, de acuerdo con la explicación más sencilla, una entidad es compleja y autónoma, dicha entidad es

real.»

La teoría de la complejidad del cálculo es la rama de la informática que se ocupa de qué recursos (tales como tiempo, capacidad de memoria o energía) son necesarios para realizar determinadas clases de cálculos. La complejidad de una información se define en términos de los recursos informáticos (tales como la extensión del programa, el número de pasos de programa o la cantidad de memoria) que necesitaría un ordenador si tuviese que reproducir dicha información. Hay varias definiciones de la complejidad, cada una de ellas con su propio campo de aplicación. No necesitamos ocuparnos aquí de los detalles de esas definiciones, pero, en general, todas se basan en la idea de que un proceso complejo es aquel que nos presenta de modo real los resultados de un cálculo sustancial. El sentido en que el movimiento de los planetas «nos presenta los resultados de un cálculo sustancial» queda bien ilustrado por un planetario. Consideremos un planetario controlado por un ordenador que calcule la imagen exacta que deben mostrar los proyectores para representar el cielo nocturno. Para hacerlo de un modo que parezca auténtico, el ordenador deberá utilizar las fórmulas proporcionadas por las teorías astronómicas. De hecho, el cálculo será idéntico al que debería realizar si estuviese calculando las predicciones para enfocar los telescopios de un observatorio a fin de localizar estrellas y planetas reales. Lo que queremos decir al afirmar que la apariencia del planetario es «tan compleja» como la del cielo nocturno que muestra, es que ambos cálculos, el que describe el cielo nocturno y el que describe el planetario, son prácticamente idénticos. Podemos así reformular de nuevo el criterio del doctor Johnson, esta vez en términos de hipotéticos cálculos: «Si es necesaria una cantidad sustancial de cálculos para proporcionarnos la ilusión de que determinada entidad es real, dicha entidad es real.»

Si el pie del doctor Johnson rebotase invariablemente cada vez que extendiese la pierna, la fuente de sus ilusiones (Dios, una máquina de realidad virtual o lo que sea) necesitaría realizar únicamente un sencillo cálculo para determinar en qué momento debería provocar la sensación de rebote (algo así como «SI pierna extendida, ENTONCES rebote...»). Pero a fin de reproducir en un experimento realista lo que sentía el doctor Johnson sería necesario tomar en consideración dónde se encuentra la piedra, sí el pie del doctor Johnson acertará a darle o no, su peso, su dureza,

con cuánta firmeza está adherida al suelo, si alguien la ha golpeado con anterioridad, etcétera, Es decir, un cálculo muy complejo.

Los físicos que procuran aferrarse a la idea de un único universo, a menudo intentan explicar los fenómenos de interferencia cuántica como sigue: «No existen los fotones fantasmas.» «No hay nada que provoque en los fotones que vemos el efecto de las ranuras distantes. Simplemente, alguna clase de acción a distancia (como en la ley de la gravedad de Newton) hace que los fotones cambien de dirección cuando es abierta una ranura distante.» Pero, en realidad, no hay nada «simple» en esa supuesta acción a distancia. La correspondiente ley física debería postular que un fotón es afectado por objetos distantes exactamente *como si* algo pasase por las ranuras distantes y rebotase en los espejos distantes para interceptar al fotón en el tiempo y el lugar adecuados. Calcular la reacción de un fotón ante esos objetos distantes requeriría el mismo esfuerzo que trabajar con la historia de grandes números de fotones fantasmas. El cálculo debería adentrarse en la historia de lo que hace cada fotón sombra: rebota aquí, es detenido por esto, y así sucesivamente. En consecuencia, al igual que con la piedra del doctor Johnson y con los planetas de Galileo, en cualquier explicación de los efectos observados aparece necesariamente una historia que trate en serio de los fotones fantasmas. La irreductible complejidad de semejante historia hace filosóficamente insostenible la negación de la existencia de los objetos.

El físico David Bohm ha elaborado una teoría, con predicciones idénticas a las de la teoría cuántica, en la que una especie de onda acompaña a cada fotón, cubre por completo la barrera, pasa por las ranuras e interfiere el fotón que vemos. La teoría de Bohm es presentada a menudo como una variante de universo único de la teoría cuántica. Pero, según el criterio del doctor Johnson, esto es un error. Investigar lo que hace la onda invisible de Bohm requiere los mismos cálculos que investigar lo que harían billones de fotones fantasmas. Algunas partes de la onda nos describen a nosotros, los observadores, cuando detectamos los fotones y reaccionamos ante ellos; otras partes describen otras versiones de nosotros que reaccionan ante los fotones en distintas posiciones. La modesta nomenclatura de Bohm —al referirse a la mayor parte de la realidad como una «onda»— no altera el hecho de que en su teoría la realidad

consiste en grandes grupos de complejas entidades, cada uno de los cuales puede percibir directamente las entidades de su propio grupo, pero sólo de modo indirecto las de los demás. Estos grupos de entidades son, en otras palabras, universos paralelos.

He descrito la nueva concepción de Galileo de nuestra relación con la realidad exterior como un gran descubrimiento metodológico. Nos proporcionó un modo de razonar nuevo y viable, que incluye la evidencia fruto de la observación. Éste es, sin duda, un aspecto fundamental de su descubrimiento: el razonamiento científico es fiable, no en el sentido de certificar que una determinada teoría permanecerá invariable, lo que no puede asegurar ni siquiera hasta mañana, sino en el de darnos la razón al confiar en él, porque hacemos lo correcto al buscar soluciones a los problemas en vez de fuentes de justificación última. La evidencia fruto de la observación es, sin duda, evidencia, no en el sentido de que cualquier teoría puede se deducida, inducida o inferida de cualquier otro modo de ella, sino en el de que puede constituir una razón genuina para preferir una teoría a otra.

Existe, sin embargo, otro aspecto del descubrimiento de Galileo mucho menos apreciado. La fiabilidad del razonamiento científico no es únicamente un atributo *nuestro*, de *nuestro* conocimiento de la realidad y *nuestra* relación con ella. Es también un hecho nuevo acerca de la propia realidad física, un hecho que Galileo expresó con la frase «el Libro de la Naturaleza está escrito en símbolos matemáticos». Como ya he dicho, es literalmente imposible «leer» ningún fragmento de una teoría en la naturaleza: éste es el error inductivista. Pero lo que sí está allí es la evidencia, o, más precisamente, una realidad que responderá con la evidencia si interactuamos adecuadamente con ella. Dado un fragmento de una teoría, o mejor dicho, fragmentos de varias teorías rivales, en la naturaleza está la evidencia que nos permitirá hacer distinciones entre ellas. Cualquiera que se tome la molestia puede buscarla, encontrarla y perfeccionarse gracias a ella.

No le harán falta autorización, iniciación ni libros sagrados. Sólo tendrá que mirar de la forma adecuada: con problemas fértiles y teorías prometedoras en la mente. Esta amplia accesibilidad, no sólo de la evidencia, sino del entero mecanismo de adquisición de conocimiento, es un atributo clave de la concepción de Galileo de la realidad.

Quizás a Galileo esto le parecía evidente de por sí, pero no lo es. Es, en realidad, una afirmación fundamental acerca de cómo es

la realidad física. Desde un punto de vista lógico, la realidad no necesita esta propiedad de ser ampliamente accesible a la ciencia, pero la tiene, y a raudales. El universo de Galileo está saturado de evidencia. Copérnico recopiló evidencia para su teoría heliocéntrica en Polonia, Tycho Brahe, en Dinamarca, y Kepler, en Alemania. Y al apuntar su telescopio al cielo de Italia, Galileo consiguió un acceso aún mayor a esa misma evidencia. Cada porción de la superficie de la Tierra, en todas las noches claras, durante millones de años, ha estado inundada de evidencia sobre los hechos y leyes de la astronomía. Y para otras muchas ciencias la evidencia ha estado expuesta de modo similar, y ha podido ser vista con mayor claridad en los tiempos modernos gracias a los microscopios y otros instrumentos. Allí donde la evidencia no está aún físicamente presente, podemos sacarla a la luz con ingenios tales como el láser y las barreras perforadas, ingenios cuya construcción es accesible a todos, en todo momento y en cualquier lugar. La evidencia será siempre la misma, sea quien sea quien la descubra. Cuando más fundamental es una teoría, más fácilmente accesible es la evidencia que la sustenta (para los que saben dónde y cómo buscarla), no sólo en la Tierra, sino en todo el multiverso.

La realidad física es *autosemejante* en varios niveles: no obstante las impresionantes complejidades del universo y el multiverso, algunas conformaciones se repiten constantemente. La Tierra y Júpiter son planetas que difieren de modo espectacular en muchos aspectos, pero ambos se desplazan en órbitas elípticas y están compuestos por el mismo grupo de aproximadamente un centenar de elementos químicos (si bien en distintas proporciones), al igual que sus contrapartidas de los universos paralelos. La evidencia que tanto impresionó a Galileo y a sus contemporáneos también existe en otros planetas y en lejanas galaxias. La evidencia que está siendo estudiada en la actualidad por físicos y astrónomos, ya estaba disponible hace miles de millones de años, y lo seguirá estando dentro de otros tantos. La propia existencia de teorías generales y explicativas implica que objetos y acontecimientos dispares son físicamente semejantes en algunos aspectos. La luz que nos llega de lejanas galaxias no es, después de todo, más que luz que nos llega con el aspecto de lejanas galaxias. La realidad, por tanto, no sólo contiene evidencia, sino también los medios (como nuestras mentes o nuestros artilugios) para comprenderla. *Hay* símbolos matemáticos en la realidad física. El hecho de que seamos nosotros quienes los

ponemos allí no hace que esos símbolos sean menos físicos. En dichos símbolos —en nuestros planetarios, nuestros libros, nuestras películas, las memorias de nuestros ordenadores o nuestros cerebros— hay innumerables imágenes de la realidad física, imágenes no sólo de la apariencia de los objetos, sino de la estructura de la realidad. Hay leyes y explicaciones, tanto reductoras como emergentes. Hay descripciones y explicaciones sobre el Big Bang, así como sobre las partículas y procesos subnucleares; hay abstracciones matemáticas, novelas, arte, moralidad, fotones fantasmas y universos paralelos. En la medida en que estos símbolos, imágenes y teorías son ciertos—es decir, se parecen de la manera apropiada a las cosas abstractas o concretas a las que se refieren—, su existencia da a la realidad una nueva clase de autosemejanza, la autosemejanza que denominamos conocimiento.

TERMINOLOGÍA

Teoría heliocéntrica. Teoría según la cual la Tierra se mueve alrededor del Sol y gira sobre su propio eje.

Teoría geocéntrica. Teoría según la cual la Tierra permanece inmóvil y los demás cuerpos astronómicos se mueven a su alrededor. Realismo. Teoría según la cual existe objetivamente un universo físico exterior que nos afecta a través de nuestros sentidos.

Navaja de Occam. Según mi formulación: *No complicar las explicaciones más allá de lo necesario*, porque, si lo hacemos, las complicaciones innecesarias quedarán sin explicar.

Criterio del doctor Johnson. Según mi formulación: *Si puede devolver el golpe, existe*. Una versión más elaborada sería: *Si, de acuerdo con la explicación más sencilla, una entidad es compleja y autónoma, dicha entidad es real*.

Autosemejanza. Propiedad según la cual ciertas partes de la realidad física (como los símbolos, las imágenes o los pensamientos humanos) se parecen a otras partes de esa misma realidad física. La semejanza puede ser concreta, por ejemplo, cuando las imágenes de un planetario imitan el cielo nocturno, o —

más importante— abstracta, por ejemplo, cuando un enunciado acerca de la teoría cuántica impreso en un libro explica correctamente un aspecto de la estructura del multiverso. (Algunos lectores tal vez estén familiarizados con la geometría de los fractales. La noción de autosemejanza definida aquí es mucho más amplia que la utilizada en dicho campo.)

Teoría de la complejidad. Rama de la informática que se ocupa de los recursos (como tiempo, capacidad de memoria o energía) necesarios para realizar determinadas clases de cálculos.

SUMARIO

Si bien el solipsismo y las teorías asociadas a él son autoconsistentes desde el punto de vista lógico, pueden ser refutadas, de modo general, simplemente, analizándolas en serio en cuanto explicaciones del mundo. Aunque todas ellas pretenden ser visiones simplificadas del mundo, dicho análisis demuestra que, en realidad, son reelaboraciones exageradas e indefendibles del realismo. Las entidades reales se comportan de modo complejo y autónomo, lo que puede ser tomado como el criterio de realidad: si algo «devuelve el golpe», existe. El razonamiento científico, que no utiliza la observación como base para la extrapolación, sino para seleccionar entre explicaciones que parecen igualmente válidas, nos puede proporcionar un genuino conocimiento sobre la realidad.

La ciencia, y otras formas de conocimiento, resultan posibles gracias a una especial propiedad de autosemejanza del mundo físico. No fueron, sin embargo, los físicos los primeros en reconocer y estudiar esta propiedad, sino los matemáticos y los teóricos de la informática, que la denominaron universalidad del cálculo. *La teoría de la calculabilidad será nuestra tercera vía*.

5. LA REALIDAD VIRTUAL

La teoría de la calculabilidad ha sido estudiada tradicionalmente en abstracto, como una rama de la matemática pura. Ello equivale a pasar por alto su principal utilidad. Los ordenadores son objetos físicos, y el cálculo es un proceso físico. Lo que un ordenador puede calcular o no, está determinado por las leyes de la física, no de la matemática pura. Uno de los conceptos más importantes de la teoría de la calculabilidad es el de *universalidad*. La definición habitual de *ordenador universal* es la de que se trata de una máquina abstracta capaz de imitar los cálculos de cualquier otra máquina abstracta dentro de una clase determinada y bien definida. Sin embargo, el significado de la universalidad reside en el hecho de que los ordenadores universales, o, al menos, aproximaciones aceptables de ellos, pueden realmente ser construidos y utilizados para calcular no sólo el comportamiento de otras máquinas de su misma clase, sino también el de entidades físicas o abstractas interesantes. El hecho de que ello sea posible es consecuencia de la autosemejanza de la realidad física de la que hablé en el capítulo anterior.

La manifestación física mejor conocida de la universalidad es una rama de la tecnología que ha sido debatida durante décadas, pero que no ha empezado a progresar hasta una época relativamente reciente: la *realidad virtual*. El término se refiere a cualquier situación en que una persona pasa de modo artificial por la experiencia de encontrarse en un entorno específico. Un simulador de vuelo —la máquina que proporciona a los pilotos la experiencia de volar sin levantarse del suelo—, por ejemplo, es un generador de realidad virtual. Esta máquina (o, más exactamente, el ordenador que la controla), puede ser programada con las características de una aeronave real o imaginaria. El entorno de la aeronave, las condiciones atmosféricas o la disposición de las pistas en los aeropuertos, por ejemplo, puede ser igualmente especificado en el programa. Mientras el piloto practica el vuelo de un aeropuerto a otro, el simulador hace aparecer las oportunas imágenes en las ventanillas, provoca las adecuadas sensaciones de sacudidas y aceleraciones, muestra las correspondientes lecturas en el panel de instrumentos, etcétera. Puede, por ejemplo, introducir los efectos de turbulencias y fallos mecánicos, así como hacer modificaciones en la aeronave. Un simulador de vuelo

puede, pues, proporcionar al usuario una amplia gama de experiencias de pilotaje, incluyendo algunas que no podría ofrecer ninguna aeronave real. La aeronave simulada podría incluso tener características no sujetas a las leyes de la física, tales como volar a través de las montañas, más deprisa que la luz o sin combustible.

Puesto que percibimos lo que nos rodea mediante nuestros sentidos, todo generador de realidad virtual debe ser capaz de manipularlos modificando su funcionamiento normal, de modo que percibamos el entorno especificado y no el real. Esto puede sonar como sacado de la novela de Aldous Huxley *Un mundo feliz*, pero, en realidad, las tecnologías para el control artificial de la experiencia sensorial humana existen desde hace milenios. Puede considerarse que todas las técnicas del arte figurativo y la comunicación a larga distancia «modifican el funcionamiento normal de los sentidos». Incluso las pinturas rupestres prehistóricas proporcionaban al observador la falsa experiencia de ver a unos animales que, en realidad, no estaban allí. Hoy podemos hacerlo con mayor precisión ayudados por películas y bandas sonoras, aunque todavía sin la suficiente perfección para que el entorno simulado se tome por el original.

Utilizaré el término *generador de imagen* para designar cualquier artefacto —como un planetario, una cadena de alta fidelidad o un especiero— capaz de provocar estímulos sensoriales especificables por el usuario: imágenes, sonidos, olores y otros estímulos específicos, todos los cuales serán considerados «imágenes». Para generar, por ejemplo, la imagen olfativa (es decir, el olor) de vainilla, abriremos el frasco de vainilla del especiero. Para generar la imagen auditiva (es decir, el sonido) del *Concierto para piano número 20* de Mozart, pondremos el correspondiente disco compacto en la cadena de alta fidelidad. Cualquier generador de imagen es un rudimentario generador de realidad virtual, aunque este término se suele reservar para los casos en que se dan simultáneamente una amplia cobertura del espectro sensorial del usuario y una intensa interacción («devolver el golpe») entre éste y las entidades simuladas.

Los videojuegos actuales permiten la interacción entre el jugador y los juegos, pero habitualmente sólo interviene una pequeña parte del espectro sensorial del usuario. El «entorno» proporcionado consiste tan sólo en imágenes en una pequeña pantalla y algunos sonidos. Ya existen, sin embargo, algunos

videojuegos en realidad virtual merecedores de tal calificativo. Típicamente, el jugador utiliza un casco con auriculares incorporados, una pantalla de televisión para cada ojo y, a veces, diversas prendas, así como guantes especiales, dotados de efectores (sistemas de generación de presión) eléctricos. Hay también sensores que detectan el movimiento de partes del cuerpo del usuario, en especial la cabeza. La información acerca de lo que hace el usuario es transmitida a un ordenador, que calcula lo que el jugador debe de estar viendo, oyendo y sintiendo en cada momento y responde enviando las correspondientes instrucciones a los generadores de imágenes (figura 5.1). Cuando el usuario mira a derecha o izquierda, las imágenes en las dos pantallas de televisión se desplazan, como haría un campo de visión real, para mostrar lo que se encuentra a su derecha o izquierda en el mundo simulado. El usuario puede alargar la mano y coger un objeto simulado, que notará como real porque los efectores situados en el guante generarán la correspondiente «retroalimentación táctil», adecuada a cualquier posición y orientación en que se perciba el objeto.

Diversos juegos y la simulación de vehículos son los principales usos de la realidad virtual actualmente, pero todo augura que su variedad aumentará de una manera espectacular en un futuro próximo. Pronto será corriente entre los arquitectos crear prototipos en realidad virtual de sus edificios, por los que sus clientes podrán deambular e introducir modificaciones en una fase en que se puede hacer con relativamente poco esfuerzo. Los compradores podrán andar (o incluso volar) por supermercados virtuales sin tener que salir de casa ni soportar aglomeraciones o una música ambiental inaguantable. El cliente no tendrá por qué sentirse solo en el supermercado virtual, ya que todas las personas que quieran podrán ir juntas de compras en realidad virtual, y cada una de ellas recibirá imágenes de las otras, al mismo tiempo que del supermercado, sin que ninguna haya tenido que dejar su domicilio. Podrán darse conciertos y conferencias sin

aglomeraciones. No sólo se ahorrará dinero en el coste de los auditorios, las entradas y los desplazamientos, sino que se dará también la ventaja de que todos los asistentes podrán ocupar simultáneamente las mejores localidades.

Si el obispo Berkeley o la Inquisición hubiesen conocido la realidad virtual, la habrían utilizado, sin duda, como perfecta ilustración de lo engañosos que son nuestros sentidos y para reforzar sus argumentos en contra del razonamiento científico. ¿Qué sucedería si el piloto de un simulador de vuelo intentase aplicar la prueba del doctor Johnson para demostrar la existencia de la realidad? Si bien el aparato simulado y su entorno no existen realmente, «devuelven el golpe» al piloto igual que si existiesen. El piloto puede acelerar y escuchar el rugido de los motores como respuesta, sentir su empuje a través del asiento y verlos vibrar y soltar gases si mira por las ventanillas, a pesar del hecho de que no interviene ningún motor. Lo único que hay fuera de la cabina es un ordenador, algunos componentes hidráulicos, pantallas de televisión, altavoces y una habitación totalmente seca y estática.

¿Invalida esto la refutación que hizo el doctor Johnson del solipsismo? No. Su conversación con Boswell habría podido tener lugar perfectamente en el interior de la cabina del simulador de vuelo. «La refuto así», podría haber dicho, al tiempo que aceleraba y notaba cómo los motores simulados «devolvían el golpe». Pero, en realidad, no hay motor alguno. Lo que «devuelve el golpe» es, en última instancia, un ordenador que ejecuta un programa, el cual calcula lo que harían los motores si fuesen «golpeados». Pero estos cálculos, exteriores a la mente del doctor Johnson, responden a la orden del acelerador del mismo modo, complejo y autónomo, en que lo harían los motores. Superan, por lo tanto, la prueba de realidad, y es muy lógico que así sea, puesto que, de hecho, esos cálculos son procesos físicos en el interior del ordenador, y éste es un objeto físico corriente —al igual que un motor— y perfectamente real. El hecho de que no haya un *motor* real es irrelevante para el argumento contra el solipsis—mo. Al fin y al cabo, no todo lo que es real debe ser identificable con facilidad. No habría importado, en la demostración original del doctor Johnson, que lo que en principio parecía una piedra hubiese resultado ser un animal camuflado o una proyección holográfica tras la que se ocultara un gnomo del jardín. Mientras su respuesta hubiese sido compleja y autónoma, el doctor Johnson habría tenido razón al concluir que era originada por algo real, externo a

él, y que, por tanto, la realidad no consistía únicamente en él.

Sin embargo, la factibilidad de la realidad virtual puede resultar un hecho incómodo para aquellas personas cuya concepción del mundo se basa en la ciencia. Parémonos a considerar qué es un generador de realidad virtual desde el punto de vista de la física. Es, por supuesto, un objeto físico, sujeto a las mismas leyes físicas que cualquier otro objeto, pero que puede «pretender» no estarlo. Puede pretender que es un objeto completamente distinto, que obedece a unas leyes espurias de la física. Y, lo que es más, puede pretenderlo de un modo complejo y autónomo. Cuando el usuario lo «golpea» para comprobar la realidad de lo que pretende ser, responde como si de veras fuese ese otro objeto inexistente y como si las leyes espurias fuesen ciertas. Si sólo dispusiéramos de semejantes objetos para aprender física, aprenderíamos leyes incorrectas. (¿O no? Por sorprendente que parezca, las cosas no están tan claras. Volveremos sobre esta cuestión en el próximo capítulo, pero primero debemos considerar el fenómeno de la realidad virtual con más detenimiento.)

Superficialmente, el obispo Berkeley parecería tener razón: la realidad virtual constituye una prueba de lo engañosos que son nuestros sentidos, y su misma factibilidad debería ponernos en guardia acerca de las innatas limitaciones de los seres humanos para la comprensión del mundo físico. Los productos de la realidad virtual parecerían deber ser incluidos, por tanto, en la misma categoría filosófica que las ilusiones, los falsos caminos y las coincidencias, ya que también estos fenómenos nos muestran algo supuestamente real, pero, en último término, nos engañan. Hemos visto que la concepción científica del mundo puede aceptar —de hecho, lo espera— la existencia de fenómenos altamente engañosos. Es, por excelencia, la concepción del mundo capaz de aceptar tanto el error humano como las fuentes externas de error. Sin embargo, los fenómenos engañosos no son exactamente bienvenidos. Aunque a veces tienen cierto valor como curiosidad, o nos muestran dónde nos hemos desviado del buen camino, tendemos a evitarlos. La realidad virtual no está, sin embargo, incluida en esa categoría de fenómenos. Veremos que la existencia de la realidad virtual no indica que la capacidad humana para la comprensión del mundo esté limitada de un modo innato, sino que, por el contrario, intrínsecamente es ilimitada. No se trata de una anomalía puesta en evidencia por las propiedades accidentales

de los órganos sensoriales humanos, sino de una propiedad fundamental del multiverso en general. Y el hecho de que el multiverso tenga esta propiedad, lejos de constituir una molesta pejiguera para el realismo y la ciencia, es esencial para ambos: es la propiedad fundamental que hace posible la ciencia. Se trata de algo de lo que, literalmente, no podríamos prescindir.

Quizás parezca un tanto exagerado dedicar semejantes alabanzas a simuladores de vuelo y videojuegos, pero es el fenómeno de la realidad virtual en general, no un determinado generador de realidad virtual, lo que ocupa un lugar primordial en el esquema de las cosas. Así pues, consideraré la realidad virtual del modo más general que sea posible. ¿Cuáles, si los tiene, son sus límites últimos? ¿Qué clases de entorno pueden ser, en principio, producidas artificialmente, y con qué fidelidad? «En principio» significa ignorar las limitaciones transitorias de la tecnología, pero tomar en consideración todas las limitaciones que puedan ser impuestas por los principios de la lógica y la física.

Tal como lo he definido, un generador de realidad virtual es una máquina que proporciona al usuario experiencias de algún entorno real o imaginario (como una aeronave) que está, o parece estar, fuera de su mente. Las denominaré *experiencias externas*. Estas experiencias deberán ser contrastadas con las *experiencias internas*, por ejemplo, los nervios al realizar el primer aterrizaje en solitario, o la sorpresa ante el súbito estallido de una tormenta en un claro cielo azul. Un generador de realidad virtual provoca en el usuario experiencias tanto externas como internas, pero no puede ser programado para hacerle pasar por una determinada experiencia interna. Por ejemplo, un piloto que realice dos veces el mismo vuelo en el simulador tendrá más o menos las mismas experiencias exteriores en ambas ocasiones, pero en la segunda, probablemente, se sorprenderá menos cuando estalle la tormenta. Por supuesto, en la segunda ocasión reaccionará también de distinto modo ante esa situación, lo que modificará, a su vez, las experiencias externas. La cuestión es que, si bien podemos programar la máquina para que haga estallar una tormenta en el campo de visión del piloto cuando queramos, no podemos, en cambio, programarla para provocar en la mente del piloto la respuesta que queramos.

Podemos concebir una tecnología, más allá de la realidad virtual, capaz de inducir también experiencias *internas* específicas. Algunas experiencias internas, tales como estados de ánimo

inducidos por ciertas drogas, pueden ser ya provocadas artificialmente, y, sin duda, el repertorio será ampliado en el futuro, pero un generador de experiencias internas específicas debería, en general, ser capaz de dominar tanto el normal funcionamiento de la mente del usuario como el de sus sentidos corporales. En otras palabras, estaría sustituyendo al usuario por otra persona. Esto sitúa a semejantes máquinas en una categoría distinta de la de los generadores de realidad virtual. Requerirán una tecnología bastante distinta y suscitarán cuestiones filosóficas de otra índole, y por ello las he excluido de mi definición de la realidad virtual.

Otra clase de experiencia que no puede, evidentemente, ser provocada de manera artificial es aquella que sea *lógicamente imposible*. He dicho que un simulador puede crear la experiencia de un vuelo a través de una montaña, algo imposible desde un punto de vista físico. Sin embargo, nada puede crear la experiencia de factorizar el número 181, puesto que ello resulta lógicamente imposible: 181 es un número primo. (*Creer* que hemos factorizado 181 es una experiencia lógicamente posible, pero a nivel interno, es decir, cae hiera del ámbito de la realidad virtual.) Otra experiencia lógicamente imposible es la inconsciencia, ya que en dicho estado, por definición, no sentimos nada. No sentir nada es algo muy distinto de experimentar una total ausencia de sensaciones —hallarse en estado de aislamiento sensorial—, que es, por supuesto, un entorno físico posible.

Tras excluir tanto las experiencias lógicamente imposibles como las internas, nos queda la amplia categoría de las *experiencias externas lógicamente posibles*, es decir, las experiencias de entornos que son posibles desde un punto de vista lógico, pero que pueden no serlo desde un punto de vista físico (véase tabla 5.1). Algo es físicamente posible cuando no lo excluyen las leyes de la física. En este libro asumiré que las «leyes de la hsica» incluyen una regla, aún desconocida, capaz de determinar el estado inicial o cualquier otro dato suplementario necesario para proporcionar, en principio, una completa descripción del multiverso (en caso contrario, estos datos constituirían un conjunto de hechos intrínsecamente inexplicables). En el caso que nos ocupa, un entorno es físicamente posible cuando —y sólo cuando— existe de manera real en algún lugar del multiverso (es decir, en algún universo o universos). Algo es físicamente imposible si no se da en ningún lugar del multiverso.

Defino el *repertorio* de un generador de realidad virtual como el conjunto de entornos, reales o imaginarios, para el cual puede ser programado a fin de proporcionar al usuario las correspondientes experiencias. Mi pregunta sobre los límites últimos de la realidad virtual puede formularse como sigue: ¿qué limitaciones, si las hay, imponen las leyes de la física a los repertorios de los generadores de realidad virtual?

La realidad virtual comprende siempre la creación de impresiones sensoriales artificiales —es decir, la generación de imágenes—, de modo que podemos empezar por aquí. ¿Qué limitaciones imponen las leyes de la física a la capacidad de los generadores de imágenes para crear imágenes artificiales, mostrarlas con todos sus detalles y cubrir sus respectivos espectros sensoriales? Hay aspectos obvios en los que un simulador de vuelo actual podría ser mejorado, por ejemplo utilizando televisores de mayor definición, pero ¿pueden llegar a ser producidos mediante la realidad virtual, aunque sólo sea en principio, un avión absolutamente real y su entorno con el máximo nivel de detalle, es decir, con el máximo nivel de detalle accesible a la capacidad de resolución de los sentidos del piloto? Por lo que hace al oído, ese nivel máximo casi ha sido alcanzado ya por las cadenas de alta fidelidad, mientras que para el de la visión se está en camino de conseguirlo. Pero ¿y para los otros sentidos? ¿Puede ser físicamente posible construir una factoría química capaz de producir al instante cualquier combinación de los millones de compuestos químicos aromáticos que existen? ¿O una máquina que, cuando se inserte en la boca de un gourmet, imite el sabor y la textura de cualquier plato? ¿Y qué decir de las sensaciones de hambre y sed que preceden a la ingesta de comida, y de las de satisfacción física que siguen a su ingestión? (El hambre, la sed y otras sensaciones, como el equilibrio y la tensión muscular, son percibidas como pertenecientes al interior del cuerpo, pero son externas a la mente y, en consecuencia, potencialmente susceptibles de entrar en el dominio de la realidad virtual.)

La dificultad para construir semejantes máquinas puede ser meramente tecnológica, pero analicemos lo siguiente: supongamos que el piloto de un simulador de vuelo dirige la aeronave simulada verticalmente hacia arriba a gran velocidad y entonces para los motores. La aeronave continuará elevándose hasta que su impulso ascendente se extinga, y entonces empezará a descender con creciente velocidad. Esta secuencia de movimientos se denomina *caída libre*, aunque la aeronave avance en sentido ascendente al principio, porque se debe exclusivamente a los efectos de la fuerza de gravedad. Cuando un aparato desciende en caída libre, sus ocupantes se hallan en estado de ingravidez y pueden flotar por la cabina igual que los astronautas cuando están en órbita. El peso no se recupera hasta que la aeronave se ve sometida de nuevo a una fuerza ascensional, lo que sucede pronto, ya sea por efecto de la aerodinámica o de la inexorable resistencia del suelo. (En la práctica, la caída libre se consigue habitualmente haciendo volar a la aeronave a potencia reducida siguiendo la misma trayectoria parabólica que describiría en ausencia simultánea de fuerza motriz y resistencia del aire.) La caída libre de las aeronaves se emplea para acostumbrar a los astronautas a los efectos de la ingravidez antes de lanzarlos al espacio. Una aeronave se mantendría en caída libre durante un par de minutos o más, puesto que podría recorrer varios kilómetros hacia arriba y luego hacia abajo. En cambio, un simulador de vuelo, que está en el suelo, sólo podrá mantenerse en esa situación unos instantes, los que tarden sus soportes en extenderse al máximo para luego empezar a retraerse. Los simuladores de vuelo (al menos los actuales) no son, pues, útiles para la habituación a la ingravidez. Para ello se necesita una aeronave.

¿Podríamos remediar esta deficiencia dando a los simuladores de vuelo la capacidad de imitar la caída libre en el suelo (en cuyo caso podrían ser también utilizados como simuladores de vuelo *espacial*)? No sería fácil, ya que las leyes de la física presentan muchas objeciones. La física conocida no ofrece otra posibilidad que la caída libre, incluso en principio, para que un objeto se halle en estado de ingravidez. El único modo de poner a un simulador de vuelo en caída libre mientras permanece estacionario sobre la superficie de la Tierra, consistiría en suspender sobre él un cuerpo masivo, por ejemplo, un planeta de masa similar a la de la Tierra o un agujero negro. Incluso si ello resultase posible (recordemos que no nos concierne aquí la

inmediata viabilidad de la propuesta, sino lo que las leyes de la física permiten o no), una aeronave, además, produciría frecuentes y complejos cambios en la magnitud y dirección del peso de sus ocupantes, debidos a las maniobras y al encendido y apagado de los motores. Para simular estos cambios, el cuerpo masivo debería poder desplazarse alrededor del simulador con la misma frecuencia, y aquí parece que, como mínimo, la velocidad de la luz establece un límite insuperable para la rapidez de dichos desplazamientos.

Sin embargo, para simular la caída libre, el simulador de vuelo no tiene por qué proporcionar una auténtica ingravidez, sino, simplemente, la experiencia de la ingravidez, y en esta dirección se han desarrollado varias técnicas que no necesitan de la caída libre para aproximarse a dicha experiencia. Los astronautas, por ejemplo, se entrenan dentro del agua vestidos con trajes espaciales lastrados de modo que tengan flotabilidad nula. Otra técnica consiste en utilizar un arnés, controlado por ordenador, que traslada al astronauta por el aire para simular la ausencia de peso. Pero se trata de métodos burdos que proporcionan sensaciones que difícilmente podrían confundirse con la realidad y mucho menos ser consideradas indistinguibles de ésta. Es inevitable que se ejerza alguna presión sobre la piel, que ésta no puede dejar de notar. Además, la característica sensación de caída, que se experimenta mediante los órganos sensoriales del oído interno, no es reproducida en absoluto. Podríamos imaginar mejoras adicionales, tales como la utilización de fluidos de muy baja viscosidad, o de drogas que provocasen la sensación de caída, pero ¿será posible llegar a proporcionar la experiencia a la perfección en un simulador de vuelo fijado firmemente al suelo? En caso contrario, habría un límite absoluto para la fidelidad de las experiencias de vuelo provocadas de manera artificial. Para distinguir entre una aeronave y un simulador, un piloto sólo debería ponerlos en trayectoria de caída libre y comprobar si experimenta o no la ingravidez.

En términos generales, el problema es el siguiente: para estimular el funcionamiento normal de los órganos sensoriales, debemos mandarles imágenes parecidas a las que produciría el entorno que se simula. Debemos también interceptar y suprimir las imágenes producidas por el entorno real del usuario. Estas manipulaciones de imágenes son operaciones físicas, y, por lo tanto, sólo pueden ser realizadas mediante procesos disponibles

en el mundo físico real. La luz y el sonido pueden ser absorbidos y reemplazados con relativa facilidad, pero, como he dicho, no ocurre lo mismo con la gravedad: las leyes de la física no lo permiten. El ejemplo de la ingravidez parece sugerir que la simulación exacta de un medio sin peso por una máquina que no vuele realmente violaría las leyes de la física.

Pero no es así. La ingravidez y las demás sensaciones pueden, en principio, ser provocadas artificialmente. Llegará el día en que sea posible obviar por completo los sentidos y estimular de manera directa los nervios que van desde ellos al cerebro.

No necesitaremos, pues, ni gigantescas fábricas químicas ni máquinas antigravitatorias de imposible realización. Cuando hayamos comprendido los órganos olfativos lo suficientemente bien para descodificar las señales que envían al cerebro cuando detectan olores, un ordenador dotado de las adecuadas conexiones con los nervios indicados podrá mandar al cerebro las mismas señales. Éste podrá entonces experimentar los olores en ausencia de los correspondientes compuestos químicos. Del mismo modo, el cerebro podrá experimentar la verdadera sensación de ingravidez, incluso en condiciones normales de gravedad. Todo ello, por supuesto, sin necesidad de auriculares ni pantallas de televisión.

Las leyes de la física no imponen, pues, limitación alguna al campo de aplicación y a la fidelidad de los generadores de imágenes. No hay ninguna sensación o secuencia de sensaciones que los seres humanos sean capaces de experimentar que no pueda, en principio, ser producida artificialmente. Algún día, como generalización de la cinematografía, habrá lo que Aldous Huxley denominaba, en *Un mundo feliz, feelies*[1], es decir, películas que permitirán poner en juego todos los sentidos. En ellas podremos notar el movimiento de un bote bajo nuestros pies, escuchar las olas y oler el mar, contemplar los cambiantes colores de la puesta de Sol en el horizonte y sentir la caricia del viento en nuestros cabellos (tanto si los tenemos como si no), y todo ello sin dejar nuestro asiento o incluso sin salir de casa. Y, lo que es más, estas películas sensoriales podrán con la misma facilidad describir escenas que nunca han existido y nunca podrán existir. Incluso podrán ofrecer, en vez de música, hermosas combinaciones abstractas de sensaciones, diseñadas para el deleite de los sentidos.

Que toda posible sensación pueda ser reproducida artificialmente es una cosa, pero que algún día sea factible

construir una máquina capaz de reproducir cualquier sensación es otra, ya que para ello se requiere algo extra: universalidad. Una máquina sensorial dotada de esta propiedad sería un *generador de imágenes universal*.

La posibilidad de un generador de imágenes universal nos obliga a cambiar de perspectiva acerca de la cuestión de los límites últimos de la tecnología sensorial. Por ahora, el progreso en esta tecnología se centra en la invención de modos más diversos y precisos de estimular los órganos sensoriales, pero esta clase de problemas desaparecerá cuando hayamos descifrado los códigos utilizados por estos órganos y desarrollemos una técnica lo bastante perfeccionada para estimular los correspondientes nervios. Una vez podamos generar señales nerviosas lo suficientemente reales para que el cerebro no note la diferencia entre ellas y las que mandarían los órganos de percepción, mejorar la precisión de esta técnica será irrelevante. Llegados a este punto, la tecnología sensorial se habrá desarrollado de tal manera que el desafío no será ya cómo proporcionar determinadas sensaciones, sino qué sensaciones proporcionar. En un ámbito concreto, esto ocurre ya, pues el problema de cómo conseguir la máxima fidelidad posible en la reproducción de sonidos ha quedado prácticamente solucionado con el disco compacto y las últimas generaciones de equipos de reproducción de sonido. Pronto no existirán los entusiastas de la alta fidelidad. El problema no será ya cuán fiel es la reproducción —siempre perfecta, norma general, hasta el límite máximo que pueda distinguir el oído humano—, sino únicamente qué valdrá la pena grabar.

Cuando un generador de imágenes reproduce una grabación tomada de la realidad, su *fidelidad* puede ser definida como el parecido entre las imágenes que ofrece y las que percibiría un observador en la situación original. De un modo más general, si el generador ofrece imágenes creadas artificialmente, como dibujos animados, o música, la fidelidad será el parecido entre las imágenes ofrecidas y las que se esperaba ver u oír. Por «parecido» entenderemos el que perciba el usuario. Si se ofrece una imagen tan parecida a la que se esperaba ver u oír que el usuario no puede distinguir la diferencia, la denominaremos *perfectamente fiel*. (En consecuencia, una representación perfectamente fiel para un usuario podrá no serlo para otro dotado de sentidos adicionales o más agudos.)

Un generador universal de imágenes no contiene, por

supuesto, las grabaciones de todas las imágenes posibles. Lo que lo hace universal es que, dada la grabación de cualquier imagen posible, puede provocar la correspondiente sensación en el usuario. Con un generador universal de sensaciones auditivas —la cadena de alta fidelidad más perfecta—, la grabación podría tener la forma de un disco compacto. Para obtener sensaciones auditivas que duren más tiempo del que permite la capacidad de un solo disco, nuestro generador deberá estar provisto de un sistema de alimentación que vaya suministrando discos a la máquina. La misma premisa es de aplicación para cualquier otro generador universal de imagen, ya que no podrá ser considerado como tal el que no sea capaz de reproducir grabaciones de manera ilimitada. Por otra parte, tras un largo tiempo de reproducción, la máquina necesitará ajustes y revisiones, pues, de lo contrario, las imágenes por ella generadas se degradarían o incluso cesarían. Estas observaciones, y otras por el estilo, están relacionadas con el hecho de que considerar que un objeto es único y está aislado del resto del universo no es más que una manera de hablar. Un generador universal de imagen sólo es universal dentro de un determinado contexto externo, en el que se supone que se le suministra energía, adecuada ventilación y mantenimiento periódico. Que una máquina tenga tales necesidades externas no la descalifica para ser considerada una «máquina universal única», siempre y cuando las leyes de la física no impidan la satisfacción de dichas necesidades y esta satisfacción no exija modificar su diseño.

Ahora bien, como he dicho, la generación de imágenes es sólo una parte de la realidad virtual: hay también un componente interactivo, no menos importante. Un generador de realidad virtual puede ser considerado como un generador de imágenes que no están completamente especificadas de antemano, sino que dependen, en parte, de lo que el usuario decida hacer. No proporciona a quien lo utiliza una secuencia de imágenes predeterminada, como haría una película, en especial, una película sensorial. Forma las imágenes a medida que avanza y toma en consideración un flujo continuo de información acerca de lo que hace el usuario. Los actuales generadores de realidad virtual, por ejemplo, siguen la posición de la cabeza del usuario mediante sensores de movimiento, como muestra la figura 5.1. En última instancia, deberán seguir todo aquello que pueda hacer el usuario que sea susceptible de afectar a la apariencia subjetiva del entorno

simulado. Este entorno puede incluir incluso el cuerpo del usuario: puesto que el cuerpo es externo para la mente, la especificación de un entorno en realidad virtual puede incluir legítimamente la exigencia de que el cuerpo del usuario parezca haber sido sustituido por otro dotado de propiedades específicas.

La mente humana afecta al cuerpo y al mundo exterior mediante la emisión de impulsos nerviosos. Por lo tanto, un generador de realidad virtual puede, en principio, obtener toda la información que necesite acerca de lo que está haciendo el usuario interceptando las señales nerviosas procedentes de su cerebro. Estas señales, en vez de llegar al cuerpo del usuario, serán transmitidas a un ordenador y descodificadas para determinar con exactitud cómo habría reaccionado dicho cuerpo. Las señales devueltas por el ordenador al cerebro pueden ser las mismas que habría enviado el cuerpo, de haberse encontrado en el entorno específico. Si este entorno así lo requiriese, el cuerpo simulado podría reaccionar de manera distinta que el real; por ejemplo, para permitirle sobrevivir en simulaciones de entornos mortales para un cuerpo real, o para simular que lo afectan enfermedades o deformidades.

Debo admitir, al llegar aquí, que quizá sea una idealización demasiado grande pretender que la mente humana *únicamente* interactúa con el mundo exterior emitiendo y recibiendo impulsos nerviosos. Circulan también mensajes químicos en ambos sentidos. Doy por sentado que, en principio, esos mensajes podrían ser igualmente interceptados y sustituidos en algún punto entre el cerebro y el resto del cuerpo. Así, el usuario, acostado e inmóvil, estaría conectado al ordenador y sentiría que interactuaba plenamente con un mundo simulado —en realidad, viviría en él—, como ilustra la figura 5.2. Incidentalmente, aunque semejante tecnología pertenezca, de momento, al futuro, la idea básica que la inspira es mucho más antigua que la teoría de la calculabilidad. A principios del siglo XVII, Descartes consideraba ya las implicaciones filosóficas de un «demonio» manipulador de los sentidos que, en esencia, no era otra cosa que un generador de realidad virtual como el mostrado en la figura 5.2, en el que una mente sobrenatural ocupaba el lugar del ordenador.

De lo que acabo de exponer parece desprenderse que todo generador de realidad virtual debería constar, al menos, de tres componentes principales:

— un conjunto de sensores (que podrían ser detectores de impulsos nerviosos), para detectar lo que hace el usuario;

— un conjunto de generadores de imágenes (que podrían ser estimuladores de los nervios), y

— un ordenador que se encargaría del control.

Mi exposición se ha concentrado hasta aquí en los dos primeros componentes: los sensores y los generadores de imágenes. Ello se debe a que en el estado actual de la tecnología sensorial, todavía incipiente, la investigación de la realidad virtual se ocupa, sobre todo, de la generación de imágenes. Pero, si miramos más allá de las limitaciones tecnológicas transitorias, veremos que los generadores de imágenes proporcionan, simplemente, la interfaz —el «cable de conexión»—entre el usuario y el verdadero generador de imágenes, que no es otro que el ordenador. El entorno es simulado por completo en el ordenador. Éste es el que proporciona el complejo y autónomo «devolver el golpe» que justifica la palabra «realidad», al hablar de «realidad virtual». El cable de conexión no aporta nada a la percepción del entorno por el usuario, y, desde el punto de vista de éste, es un elemento «transparente», del mismo modo que no percibimos nuestros nervios como parte de nuestro entorno. Así pues, una adecuada descripción de los generadores de realidad virtual del futuro es que estarán constituidos por un solo elemento principal —el ordenador—, más algunos artilugios periféricos triviales.

No es mi intención subestimar los problemas prácticos involucrados en la interceptación de todas las señales nerviosas que entran y salen del cerebro humano, así como en descifrar los correspondientes códigos, pero éste es un conjunto de dificultades que deberán ser resueltas sólo una vez. Después de ello, la atención de la tecnología de la realidad virtual se centrará exclusivamente en el ordenador, en el problema de programarlo para reproducir diversos entornos. Cuáles seamos capaces de reproducir no dependerá ya de qué sensores y qué reproductores de imágenes seamos capaces de construir, sino de qué entornos podamos especificar. «Especificar» un entorno significará

suministrar un programa al ordenador, verdadero corazón del generador de realidad virtual.

El concepto de reproducción fiel tiene más importancia para la generación de imágenes que para la realidad virtual, a causa de la naturaleza interactiva de ésta. Como he dicho, la fidelidad de un generador de imágenes se mide por la similitud entre las imágenes obtenidas y las que se esperaba obtener. Pero en la realidad vidual, normalmente, no se espera obtener *imágenes*; lo que se intenta conseguir es un determinado entorno para que lo experimente el usuario. Especificar un entorno de realidad virtual no significa especificar lo que el usuario va a experimentar, sino más bien el modo en que dicho entorno responderá a cada una de las posibles acciones de aquél. En un partido simulado de tenis, por ejemplo, se puede especificar de antemano el aspecto de la pista, el tiempo que hará, el comportamiento del público y la categoría como jugador del oponente, pero no cómo se desarrollará el partido, lo cual dependerá del conjunto de las decisiones que tome el usuario mientras juega. Cada conjunto de decisiones tendrá como resultado diferentes respuestas por parte del entorno simulado y, en consecuencia, un partido distinto.

El número de posibles partidos que se pueden jugar en un mismo entorno —es decir, producidos por un mismo programa— es enorme. Consideremos una reproducción de la pista central de Wimbledon desde el punto de vista del jugador. Supongamos, muy conservadoramente, que en cada segundo del partido el usuario puede actuar de una o dos maneras perceptibles (es decir, perceptibles para él). Luego, tras dos segundos de juego hay cuatro acciones posibles, tras tres, ocho, y así sucesivamente. Tras unos cuatro minutos, el número de acciones posibles, perceptiblemente distintas entre sí, excede el número de átomos del universo, y sigue aumentando de modo exponencial. Para que un programa pueda reproducir con fidelidad un entorno así, debe ser capaz de responder a cada una de esa infinidad de posibilidades perceptiblemente distintas en función de las decisiones que adopte el jugador. Si dos programas responden del mismo modo a cada posible acción del usuario, producen el mismo entorno. Si reaccionan de manera perceptiblemente distinta, incluso ante una sola acción, producen entornos distintos.

Esto será de aplicación aun cuando el usuario no realice jamás la acción que provoca la diferencia. El entorno que ofrece un programa (para un tipo determinado de usuario, con un tipo

determinado de cable de conexión) es una propiedad lógica de ese programa, con independencia de que sea ejecutado o no. Un entorno reproducido es fiel en la medida en que *respondería* del modo previsto a cada posible acción del usuario. Esta fidelidad dependerá, pues, no sólo de las experiencias que el usuario obtenga realmente, sino también de las que hubiese podido obtener de haber decidido actuar de modo distinto durante la ejecución. Esto puede sonar paradójico, pero, como he dicho, es consecuencia directa del hecho de que la realidad virtual es, como la propia realidad, interactiva.

Esto da pie a una importante diferencia entre la generación de imágenes y la de realidad virtual. La fidelidad de representación de un generador de imágenes puede ser, en principio, experimentada, medida y certificada por el usuario, pero la fidelidad de representación de la realidad virtual no. Si es usted, por ejemplo, melómano y conoce una pieza lo suficientemente bien, podrá escuchar su interpretación y juzgar si se trata o no de una reproducción fiel en todas sus notas, frases, dinámica y demás. Pero aunque usted sea un aficionado al tenis y conozca a la perfección la pista central de Wimble—don, nunca podrá confirmar que la reproducción que le ofrecen de ella es fiel. Aunque pudiese explorar la reproducción propuesta tanto tiempo como quisiera y «golpearla» de cualquier modo que se le ocurriera, e incluso aunque tuviese idéntico acceso a la auténtica pista para comparar, nunca podría certificar que el programa reproducía fielmente el entorno real, porque nunca sabría lo que hubiese podido suceder de haber explorado un poco más o haber mirado por encima de su hombro en el momento oportuno. Quizás, si se hubiese sentado en la silla del juez y gritado *Fault!*, habría emergido de la pista un submarino nuclear que habría torpedeado el marcador.

Pero, en cambio, si encuentra una sola diferencia entre la representación y el entorno propuesto, podrá certificar en el acto que la representación es *infiel*. Es decir, a menos que el entorno propuesto tenga algunas prestaciones intencionalmente impredecibles. Una ruleta, por ejemplo, está diseñada para ser impredecible. Si filmamos una película del funcionamiento de una ruleta en un casino, dicha película será fiel si reproduce los mismos números que salieron cuando se filmó. La película mostrará los mismos números cada vez que sea proyectada, de modo que es totalmente predecible. Así pues, una *imagen* fiel de

un entorno impredecible, debe ser predecible. Pero ¿qué significa que la *representación* de una ruleta en realidad virtual sea fiel? Al igual que antes, significa que el usuario no debería encontrar diferencia alguna perceptible con el original. Pero ello implica que la representación *no* debe comportarse de modo idéntico al original. Si lo hiciera, tanto ella como el original podrían ser utilizados para predecir sus comportamientos respectivos, con lo cual dejarían de ser impredecibles. Tampoco debe comportarse del mismo modo cada vez que funciona. Una representación perfecta de una ruleta ha de ser utilizable exactamente del mismo modo que la real, y, por lo tanto, debe ser también impredecible. Debe ser, asimismo, igual de justa, es decir, los números deben poder salir de forma totalmente aleatoria, con idénticas probabilidades.

¿Cómo reconocemos entornos impredecibles, y cómo confirmamos que unos números deliberadamente aleatorios están distribuidos de manera justa? Comprobamos que la representación de la ruleta cumple sus especificaciones, del mismo modo que comprobamos que lo hace la ruleta real, «golpeándola» (haciéndola girar) y viendo si responde según lo anunciado. Efectuamos gran número de observaciones similares y proyecciones estadísticas de los resultados. De nuevo, por más pruebas que realicemos, nunca podremos certificar que la representación sea no ya fiel, sino ni siquiera probablemente fiel. Y es que, por más aleatorio que parezca el modo en que salen los números, podrían seguir un patrón secreto que permitiría su predicción al usuario que lo conociese. O quizás la ruleta reaccionase a determinados estímulos, y cada vez que preguntásemos en voz alta la fecha de la batalla de Waterloo, saldrían invariablemente los números 18 y 15, que forman la fecha correcta. Por otra parte, si la secuencia en que aparecen los números nos parece injusta, nunca podremos estar seguros de que lo sea, sólo de que, *probablemente*, lo es. Si, por ejemplo, el número cero aparece en la representación de la ruleta en diez tiradas consecutivas, podríamos deducir que, probablemente, no disponemos de una representación fiel de la ruleta.

Al hablar de los generadores de imágenes, decíamos que la fidelidad de una imagen reproducida depende de la agudeza y demás atributos de los sentidos del usuario. Con la realidad virtual en cambio, éste es el menor de los problemas. Sin duda, un generador de realidad virtual que reproduzca un determinado entorno a la perfección para los humanos, no lo hará para los

delfines o los extraterrestres. Para reproducir un entorno determinado para un usuario dotado de cierta clase de órganos sensoriales, un generador de realidad virtual debe estar físicamente adaptado a esos órganos, y su ordenador debe haber sido programado con sus características. No obstante, las modificaciones necesarias para adaptarse a una determinada clase de usuario son finitas y sólo han de hacerse una vez. Se resumen en lo que he denominado «la construcción de un nuevo cable de conexión». Al considerar entornos de complejidad creciente, la tarea de representarlos para una determinada clase de usuario se centra, sobre todo, en el diseño de los programas necesarios para calcular lo que harán esos entornos. La parte de la tarea específica para la especie, al ser de complejidad fija, tiene, en cambio, menos importancia. La presente reflexión trata de los límites últimos de la realidad virtual, de modo que consideramos representaciones arbitrariamente fieles, largas y complejas. En este sentido, podemos hablar de «reproducir un determinado entorno» sin especificar para quién.

Hemos visto que existe una noción bien definida de fidelidad para una representación en realidad virtual: la semejanza, hasta donde sea perceptible, del entorno representado con el que se deseaba representar. Pero debe serlo para cualquier comportamiento posible del usuario, y, por esta razón, por más atención con que observemos el entorno representado, nunca podremos certificar que es fiel (o probablemente fiel). Sin embargo, la experiencia puede, en ocasiones, mostrar que una determinada representación es infiel (o probablemente infiel).

Esta discusión acerca de la fidelidad en realidad virtual ejemplifica la relación entre teoría y experimento en ciencia. Aquí también es posible confirmar experimentalmente que una teoría general es falsa, pero nunca que sea cierta. Del mismo modo, una concepción estrecha de la ciencia es la de que ésta sólo trata de la predicción de nuestras impresiones sensoriales. La concepción correcta es que, si bien las impresiones sensoriales siempre tendrán un papel, de lo que realmente trata la ciencia es de la comprensión de la totalidad de la realidad, de la que sólo llega a conocerse por experiencia propia una parte infinitesimal.

El programa de un generador de realidad virtual da forma material a una teoría predictiva y general sobre el comportamiento del entorno representado. Los demás componentes se ocupan del seguimiento de las acciones del usuario y la codificación y

descodificación de datos sensoriales, funciones, como he dicho, relativamente triviales. Así pues, si el entorno es físicamente posible, representarlo equivale, en esencia, a encontrar las reglas para predecir el resultado de cada experimento que podría ser realizado en dicho entorno. A causa del modo en que se crea el conocimiento científico, sólo mejores teorías explicativas permiten descubrir reglas de predicción cada vez más precisas. Por consiguiente, reproducir con fidelidad cualquier entorno físicamente posible depende de la comprensión de su física.

Lo contrario es también cierto: el descubrimiento de la física de un entorno depende de que se cree su reproducción en realidad virtual. A primera vista, parece que las teorías científicas describen y explican objetos y procesos físicos, pero no los representan. Una explicación de los eclipses de Sol, por ejemplo, puede ser impresa en un libro. Un ordenador podría ser programado con datos astronómicos y leyes físicas para predecir un eclipse e imprimir su descripción. Pero representar el eclipse en realidad virtual exigiría más programación y más equipo. Ahora bien, ambas cosas ya existen en nuestro cerebro. Las palabras y los números impresos por el ordenador contribuyen a una «descripción» del eclipse sólo para quien conoce el significado de estos símbolos, que evocan en la mente de quien los interpreta una especie de imagen del efecto previsto del eclipse, y contra este modelo se contrastará la versión real del fenómeno. Además, la imagen evocada es interactiva. Podemos observar un eclipse de muchas maneras: a simple vista, mediante fotografías o utilizando diversos instrumentos científicos. Desde ciertos lugares de la Tierra veremos un eclipse total, desde otros, parcial, y desde otros, el eclipse no será visible. En cada caso, el observador verá distintas imágenes, cualquiera de las cuales puede ser predicha por la teoría. Lo que la descripción proporcionada por el ordenador evoca en la mente de quien interpreta los datos, no es una sola imagen o un conjunto de imágenes, sino un método general para la creación de diferentes imágenes, correspondientes a las múltiples maneras en que aquél pueda plantearse la realización de observaciones. En otras palabras, es una reproducción en realidad virtual. Así, en un sentido muy amplio, y tomando en consideración los procesos que deben desarrollarse en la mente del científico, ciencia y representación en realidad virtual de entornos físicamente posibles son dos términos que denotan la misma actividad.

Muy bien. ¿Y qué hay de la representación de entornos

físicamente imposibles? A primera vista, hay dos clases de representación en realidad virtual: una, minoritaria, describe entornos físicamente posibles, mientras que la otra, mayoritaria, describe entornos físicamente imposibles. Pero ¿puede esta distinción superar a un examen más atento? Consideremos un generador de realidad virtual en el acto de reproducir un entorno físicamente imposible. Podría ser un simulador de vuelo ejecutando un programa que calculara la visión desde la cabina de una aeronave, que volase a velocidad superior a la de la luz. El simulador estaría *reproduciendo* este entorno. Pero, además, el simulador sería el propio entorno que notaría el usuario, en el sentido de que es un objeto físico que lo rodea. Consideremos este entorno. Es, sin duda, físicamente posible. ¿Es representable? Sí. De hecho, es facilísimo de representar: sólo necesitamos otro simulador idéntico que ejecute el mismo programa. En las circunstancias a las que nos referimos, podemos considerar que el segundo simulador estaría reproduciendo, bien la aeronave físicamente imposible, bien un entorno físicamente posible (en el caso que nos ocupa, el primer simulador). Del mismo modo, podríamos considerar que el primer simulador estaría reproduciendo un entorno físicamente posible (en este caso, el segundo simulador). Si aceptamos que cualquier generador de realidad virtual que pueda ser, en principio, construido, puede, en principio, ser construido de nuevo, se sigue que todo generador de realidad virtual que ejecuta un programa de su repertorio está reproduciendo *algún* entorno físicamente posible. Puede estar reproduciendo también otras cosas, incluyendo entornos físicamente imposibles, pero, en particular, siempre habrá algún entorno físicamente posible que esté reproduciendo.

Así pues, ¿qué entornos físicamente imposibles pueden ser reproducidos en realidad virtual? Precisamente los que no sean perceptiblemente distintos de entornos físicamente posibles. Por tanto, la conexión entre el mundo físico y los mundos representables en realidad virtual es mucho más estrecha de lo que parece. Consideramos que ciertas representaciones en realidad virtual muestran hechos reales, mientras que otras muestran hechos ficticios, pero lo ficticio es siempre una interpretación en la mente del observador. No existe ningún entorno de realidad virtual que el usuario tenga que interpretar por fuerza como físicamente imposible.

Podríamos decidir representar un entorno según las

predicciones de unas determinadas «leyes de la física» distintas de las verdaderas. Podríamos hacerlo como ejercicio, o para divertirnos, o para conseguir una aproximación lo más ajustada posible en el caso de que la representación fiel resultase demasiado difícil o cara. Si las leyes que empleásemos fueran lo más parecidas a las reales que fuese posible —dependiendo de las limitaciones con que estuviésemos operando—, podríamos denominar a dicha representación «matemática aplicada» o «cálculo». Si los objetos representados fuesen muy distintos de los físicamente posibles, hablaríamos de «matemática pura». Cuando se representa un entorno físicamente imposible como diversión, lo denominamos «videojuego» o «arte cibernético». Todo esto son interpretaciones. Pueden ser interpretaciones útiles, o incluso esenciales, a fin de explicar nuestros motivos para realizar una determinada representación, pero, por lo que a ésta concierne, existirá siempre una interpretación alternativa, a saber, la que describe con fidelidad algún entorno físicamente posible.

No es habitual considerar a las matemáticas una forma de realidad virtual. Normalmente, consideramos que tratan de entidades abstractas, como números y conjuntos, que no afectan a los sentidos, y, por consiguiente, podría parecer, en principio, que no existe manera alguna de representar artificialmente su efecto sobre nosotros. Sin embargo, aunque las entidades matemáticas no afecten a los sentidos, trabajar con ellas constituye una experiencia externa, al igual que realizar experimentos físicos. Trazamos signos sobre un papel y los miramos, o imaginamos que lo hacemos. Es evidente que no podemos realizar ninguna tarea matemática sin imaginar entidades matemáticas abstractas, por ello significa, en realidad, imaginar un entorno cuya «física» de forma material a las complejas y autónomas propiedades de dichas entidades. Cuando, por ejemplo, imaginamos el concepto abstracto de un segmento lineal carente de anchura, podemos imaginarnos una línea visible, pero imperceptiblemente ancha. Esto podría acomodarse, más o menos, a la realidad física. Pero, desde un punto de vista matemático, la línea debe seguir careciendo de ancho aunque sea ampliada de manera arbitraria. Ninguna línea física tiene esta propiedad, pero puede atribuírsela sin dificultad la realidad virtual de nuestra imaginación.

La imaginación es una forma evidente de realidad virtual. Lo que quizás no resulte tan evidente es que nuestra experiencia «directa» del mundo, a través de los sentidos, también es realidad

virtual. Nuestra experiencia externa nunca es directa. Ni siquiera experimentamos directamente las señales que recorren nuestros nervios, y, de ser posible, no sabríamos qué hacer con el torrente de chasquidos eléctricos que transportan. Lo que experimentamos de manera directa es una representación en realidad virtual, convenientemente generada para nosotros por nuestra mente inconsciente a partir de datos sensoriales y con la ayuda de complejas teorías innatas y adquiridas (es decir, programas) acerca de cómo interpretarlos.

Los realistas creemos que la realidad está en todas partes, objetiva, física e independiente de lo que pensemos de ella, pero que nunca la experimentamos directamente. Hasta la última brizna de nuestra experiencia externa es realidad virtual. Hasta la última brizna de nuestro conocimiento —incluyendo nuestro conocimiento de los mundos no físicos de la lógica, las matemáticas y la filosofía, así como de la imaginación, el arte, la ficción y la fantasía— está codificada en forma de programas para la representación de esos mundos en el generador de realidad virtual que es nuestro cerebro.

No es, pues, únicamente la ciencia —el razonamiento acerca del mundo físico— lo que se relaciona con la realidad virtual. Todo razonamiento, todo pensamiento y toda experiencia externa son formas de realidad virtual. Se trata de procesos físicos observados hasta ahora en un solo lugar del universo: el planeta Tierra. Veremos en el capítulo 8 que todos los procesos vitales se relacionan asimismo con la realidad virtual, si bien los seres humanos tenemos nna relación especial con ella. Biológicamente hablando, la representación en realidad virtual de nuestro entorno es el modo característico que tenemos los humanos para sobrevivir. Dicho de otro modo, es la razón por la que existimos. El nicho ecológico que ocupamos los seres humanos depende de la realidad virtual de manera tan directa y absoluta como el de los koalas de las hojas de eucalipto.

TERMINOLOGÍA

Generador de imágenes. Aparato capaz de provocar sensaciones es—pecificables en un usuario. Se llama *universal* cuando es susceptible de ser programado para crear cualquier sensación que el usuario pueda experimentar.

Experiencia. Uno de los procesos de adquisición de conocimiento. Se divide en *interna y externa*, según que ocurra dentro o fuera de la mente individual.

Posibilidad física. Es físicamente posible todo aquello que no contravenga las leyes de la física. Un entorno es físicamente posible cuando —y sólo cuando— existe en algún lugar del multiverso (por ello, se presume que las condiciones iniciales y demás datos suplementarios del multiverso pueden estar determinados por algunas leyes de la física aún desconocidas).

Posibilidad lógica. Es lógicamente posible todo aquello que sea auto—consistente.

Realidad virtual. Toda situación en que un usuario experimenta la sensación de encontrarse en un entorno específico.

Repertorio. En un generador de realidad virtual, conjunto de entornos para los cuales puede ser programado dicho generador a fin de proporcionar al usuario las correspondientes experiencias.

Imagen. Todo aquello que provoca sensaciones.

Fidelidad. Medida en que las sensaciones que provoca una imagen se parecen a las que se deseaba provocar. La representación de un entorno será fiel en la medida en que responda del modo en que se deseaba que lo hiciera ante cada posible acción del usuario.

Fidelidad perfecta. Fidelidad tan grande que el usuario no es capaz de distinguir la imagen o el entorno producidos de los que se deseaba producir.

SUMARIO

La realidad virtual no es tan sólo una tecnología en la que unos ordenadores simulan el comportamiento de un entorno físico.

El hecho de que la realidad virtual sea posible es un factor importante en la estructura de la realidad. Es la base no sólo del cálculo, sino de la imaginación humana y la experiencia externa, de la ciencia y las matemáticas, del arte y la ficción.

¿Cuáles son los límites últimos —el alcance completo— de la realidad virtual (y, en consecuencia, del cálculo, la ciencia, la imaginación, etcétera)? En el próximo capítulo veremos que en un aspecto el alcance de la realidad virtual es ilimitado, mientras que en otro está drásticamente circunscrito.

6. LA UNIVERSALIDAD Y LOS LÍMITES DEL CÁLCULO

El corazón de un generador de realidad virtual es su ordenador, y la pregunta de qué entornos pueden ser reproducidos en realidad virtual por fuerza ha de conducir a la de qué cálculos pueden ser realizados. El repertorio de los generadores de realidad virtual todavía está limitado tanto por sus ordenadores como por sus generadores de imágenes. Cada vez que se incorpora a un generador de realidad virtual un ordenador más potente y más rápido, así como un mejor equipo de procesamiento de imágenes, el repertorio se amplía. Pero ¿seguirá siendo así? ¿Alcanzaremos finalmente la plena universalidad, tal como ya he dicho que cabe esperar en el caso de los generadores de imágenes? En otras palabras, ¿puede llegar a construirse un generador de realidad virtual único susceptible de ser programado para reproducir cualquier entorno que la mente humana sea capaz de experimentar?

Tal como ocurre con los generadores de imágenes, no quiero decir con esto que el generador de realidad virtual único deba contener las especificaciones de *todos* los entornos lógicamente posibles, sino que pueda ser programado para reproducir cualquiera de ellos. Podemos codificar los programas, por ejemplo, en discos compactos. Cuanto más complejo sea el entorno, más discos serán necesarios para almacenar el correspondiente programa. Por lo tanto, para reproducir entornos complejos la máquina deberá disponer de un mecanismo de alimentación —como en el caso del generador universal de imágenes— capaz de leer cantidades ilimitadas de discos. A diferencia del generador de imágenes, un generador de realidad virtual puede necesitar una creciente cantidad de «memoria de trabajo» que almacene los resultados intermedios de sus cálculos. Podemos proporcionársela en forma de discos en blanco. El hecho de que la máquina necesite un suministro de energía y discos en blanco, así como mantenimiento, no nos impide considerarla una «única máquina», a condición de que esas operaciones no equivalgan a una modificación de su diseño ni contravengan las leyes de la física.

En este sentido, pues, puede ser contemplado, en principio, un ordenador con una capacidad de memoria verdaderamente ilimitada, pero no, en cambio, un ordenador con velocidad de cálculo ilimitada. Un ordenador con un determinado diseño tendrá

siempre una velocidad máxima, que sólo se podrá aumentar mediante cambios en el diseño. Así pues, un determinado generador de realidad virtual no podrá realizar cantidades ilimitadas de cálculos por unidad de tiempo. ¿No limitará esto su repertorio? Si un entorno es tan complejo que se tarda más de un segundo en calcular lo que el usuario debería ver dentro de un segundo, ¿cómo podrá la máquina reproducirlo fielmente? Para alcanzar la universalidad, necesitamos un truco tecnológico adicional.

Para extender su repertorio tanto como sea físicamente posible, un generador de realidad virtual deberá controlar otro de los atributos del sistema sensorial del usuario, a saber, la velocidad de proceso de su cerebro. Si el cerebro humano fuese igual que un ordenador electrónico, sólo sería necesario cambiar el ritmo en el que su «reloj» emite los impulsos sincronizadores. Sin duda, el «reloj» del cerebro no es tan fácilmente controlable. Sin embargo, esto tampoco presenta, en principio, problema alguno. El cerebro es un objeto físico finito, y, por lo tanto, todas sus funciones son procesos físicos que, en principio, pueden ser ralentizados o detenidos. El generador de realidad virtual definitivo deberá ser capaz de hacerlo.

Para conseguir una perfecta representación de entornos que requieran muchos cálculos, un generador de realidad virtual debería operar más o menos del modo siguiente: cada nervio sensorial es físicamente capaz de transmitir señales a un cierto ritmo máximo, ya que una célula nerviosa que acaba de transmitir no puede volverlo a hacer hasta haber transcurrido aproximadamente un milisegundo; por lo tanto, después que un nervio ha transmitido, el ordenador dispone al menos de un milisegundo para decidir si, y cuándo, dicho nervio debe transmitir de nuevo. Si ha calculado esta decisión en, digamos, medio milisegundo, no será necesaria manipulación alguna de la velocidad del cerebro y el ordenador, simplemente, conectará el nervio en los momentos adecuados. De otro modo, el ordenador deberá provocar la ralentización del cerebro (o, en caso necesario, su detención) hasta haber completado el cálculo de lo que habría de suceder a continuación, y entonces volverá a permitir funcionar al cerebro a su velocidad normal. ¿Qué sensación notaría el usuario? Por definición, ninguna. Experimentaría sólo el entorno especificado en el programa, sin ninguna ralentización, detención o reanudación. Afortunadamente, nunca es necesario que un

generador de realidad virtual acelere el normal funcionamiento del cerebro, lo que provocaría, sin duda, problemas de principio, ya que, entre otras cosas, ninguna señal puede viajar a mayor velocidad que la de la luz.

Este método nos permite especificar de antemano un entorno arbitrariamente complicado, cuya simulación requiera cualquier cantidad finita de cálculos y experimentarlo a cualquier velocidad y nivel de detalle subjetivos que nuestras mentes sean capaces de asimilar. Si los cálculos requeridos son demasiado numerosos para que los pueda realizar el ordenador, dentro del tiempo percibido subjetivamente, la experiencia no se verá afectada, si bien el usuario pagará esa complejidad en términos de tiempo externo transcurrido. Al terminar una experiencia con un generador de realidad virtual que hubiera tenido una duración subjetiva de cinco minutos, el usuario podría encontrarse con que habían transcurrido cinco años en la realidad física.

Un usuario cuyo cerebro sea desactivado y luego vuelto a activar tendrá una experiencia ininterrumpida de algún entorno, aunque haya pasado un largo período de tiempo en esa situación, pero un usuario cuyo cerebro sea desactivado para siempre dejará de tener experiencias a partir del momento de la desconexión. Esto significa que un programa que desactivara el cerebro del usuario y no lo volviera a activar, no generaría ningún entorno que pudiera ser experimentado por aquél y, por lo tanto, no podría ser considerado válido para un generador de realidad virtual. En cambio, un programa que reactive, antes o después, el cerebro del usuario obliga al generador de realidad virtual a reproducir algún entorno. Incluso un programa que no emitiese señal nerviosa alguna estaría reproduciendo algo: el oscuro y silencioso entorno del aislamiento sensorial perfecto.

En nuestra búsqueda de los límites de la realidad virtual nos hemos alejado mucho de lo que es posible hoy día e incluso de lo que nos traerá el futuro tecnológico más previsible. Permítaseme, pues, insistir en que para nuestro propósito los obstáculos tecnológicos son irrelevantes. No investigamos qué clases de generadores de realidad virtual podemos construir, ni siquiera qué clases de generadores de realidad virtual podrán ser construidos por ingenieros humanos. Investigamos lo que las leyes de la física permiten o no en el campo de la realidad virtual. Y ello es importante por una razón que no tiene nada que ver con las perspectivas de construir mejores generadores de realidad virtual.

Dicha razón es que la relación entre la realidad virtual y la realidad «ordinaria» forma parte de la compleja e insólita estructura de la realidad, que es el tema de este libro.

Al considerar diversos trucos —estimulación nerviosa, detención y reactivación del funcionamiento del cerebro, etcétera—, hemos conseguido proponer un generador de realidad virtual físicamente posible y cuyo repertorio cubre la totalidad del espectro sensorial, el cual es plenamente interactivo y no está limitado por la velocidad o la capacidad de memoria de su ordenador. ¿Queda algo fuera del repertorio de semejante generador de realidad virtual? ¿Constituiría su repertorio el catálogo de todos los entornos físicamente posibles? La respuesta es que no. Incluso el repertorio de esta máquina futurista quedaría drásticamente limitado por su mera condición de objeto físico. En realidad, ni siquiera llegaría a rozar la superficie de lo que es lógicamente posible, como demostráré a continuación.

La idea básica de la prueba que lo demuestra —conocida como *argumento diagonal*— es anterior a la idea de la realidad virtual. Fue empleada por primera vez en el siglo XIX por el matemático Georg Cantor a fin de probar la existencia de cantidades infinitas mayores que la infinidad de los números naturales (1, 2, 3, ...). La misma prueba constituye el núcleo de la teoría de la calculabilidad, desarrollada por Alan Turing y otros matemáticos en los años treinta, y fue utilizada también por Kurt Gödel para probar su conocido «teorema de la in—completitud», del que hablaremos en el capítulo 10.

Cada entorno del repertorio de nuestra máquina es generado por algún programa de su ordenador. Imaginemos el conjunto de todos los programas válidos de dicho ordenador. Desde un punto de vista físico, cada uno de esos programas especifica un determinado conjunto de valores para variables físicas —sobre discos u otro soporte—con que se programa el ordenador. Sabemos, por la teoría cuántica, que todas estas variables están cuantificadas y que, por consiguiente, con independencia de cómo trabaje el ordenador, el conjunto de posibles programas será discreto. Cada programa puede, por lo tanto, ser expresado como una secuencia finita de símbolos en un código discreto o en lenguaje informático. El número de posibles programas es infinito, pero cada uno de ellos sólo puede contener un número finito de símbolos porque los símbolos son objetos físicos, hechos de materia y con configuraciones reconocibles, y no podemos

confeccionar una cantidad infinita de ellos. Como explicaré en el capítulo 10, esas exigencias físicas intuitivamente obvias (que los programas deben estar cuantificados y cada uno de ellos ha de consistir en un número finito de símbolos y puede ser ejecutado en una secuencia de pasos) son más importantes de lo que parece. Son las únicas consecuencias de las leyes de la física que deben ser respetadas estrictamente para que sea posible la prueba, pero resultan suficientes para imponer drásticas restricciones al repertorio de cualquier máquina físicamente factible. Otras leyes físicas podrán imponer restricciones adicionales, pero éstas no afectan a las conclusiones del presente capítulo.

Imaginemos ahora ese conjunto infinito de programas posibles, recogidos en una lista infinitamente larga y numerados como programa 1, programa 2, etcétera. Podrían, por ejemplo, estar ordenados por orden «alfabético» de acuerdo con los símbolos en los que estén expresados. Puesto que cada programa genera un entorno, la lista podría también ser una relación de todos los entornos del repertorio de la máquina, que podríamos denominar entorno 1, entorno 2, etcétera. Podría ocurrir que alguno de los entornos estuviese repetido, ya que es efectivamente posible que dos programas distintos efectúen los mismos cálculos, pero ello no afecta al razonamiento. Lo importante es que cada entorno del repertorio de nuestra máquina aparezca, al menos, una vez en la lista.

Un entorno simulado puede ser limitado o ilimitado en tamaño y duración aparentes. La simulación de una vivienda por un arquitecto, por ejemplo, puede ejecutarse por tiempo indefinido, pero tendrá, en cambio, un volumen limitado. Un videojuego puede conceder al usuario un tiempo limitado para jugar, o puede reproducir un universo de juego de dimensiones ilimitadas, permitir una capacidad de exploración ilimitada y acabar sólo cuando el usuario lo decida deliberadamente. Para simplificar la prueba, consideraremos únicamente programas sin fin. No es una gran restricción, ya que si un programa se para, siempre podemos considerar su falta de respuesta como la respuesta de un entorno de aislamiento sensorial.

Permítaseme definir aquí una clase de entornos lógicamente posibles que denominaré *entornos cantgotu*, en parte en honor de *Can*tor, *Gö*del y *Tu*ring, y en parte por la razón que explicaré en breve. Durante el primer minuto subjetivo, un entorno cantgotu se comporta de modo distinto al entorno 1 (generado por el

programa 1 de nuestro generador). No importa cómo se comporte, mientras sea para el usuario reconociblemente distinto del entorno 1. Durante el segundo minuto, se comporta de modo distinto al entorno 2 (si bien ahora puede parecerse al entorno 1). Durante el tercer minuto, se comporta de modo distinto al entorno 3 y así sucesivamente. Denominaré entorno cantgotu a todo entorno que satisfaga estas reglas.

Ahora bien, puesto que un entorno cantgotu no se comporta exactamente como el entorno 1, no puede *ser* el entorno 1; y puesto que no se comporta exactamente como el entorno 2, no puede *ser* el entorno 2. Puesto que es seguro que, antes o después, se comportará de modo distinto al entorno 3, al entorno 4 y a todos los entornos de la lista, tampoco puede ser ninguno de ellos. Pero la lista contiene todos los entornos generados por todos los programas posibles para esa máquina. Por consiguiente, ninguno de los entornos cantgotu está en el repertorio de la máquina. Los entornos cantgotu son entornos a los que *no podemos acceder*[2] desde este generador de realidad virtual.

Está claro que existen infinidad de entornos cantgotu, ya que su definición deja una inmensa libertad para escoger su comportamiento; la única restricción es que, durante cada minuto, no deben comportarse de una determinada manera. Puede demostrarse que por cada entorno del repertorio de un determinado generador de realidad virtual hay infinitamente más entornos cantgotu que no puede reproducir. Tampoco hay mucho campo para ampliar el repertorio utilizando una gama distinta de generadores de realidad virtual. Supongamos que tuviésemos un centenar, cada uno de ellos dotado (para los propósitos del razonamiento) de un repertorio diferente. Toda la colección, combinada con el sistema de control programable que determina cuál de ellos debe entrar en funcionamiento para ejecutar un determinado programa, no sería más que un generador de realidad virtual más grande. Este generador estaría sujeto al razonamiento que he expuesto, así que por cada entorno que pudiese reproducir habría una infinidad que no podría. Y, lo que es más, la suposición de que diferentes generadores de realidad virtual podrían tener diferentes repertorios peca de optimista. Como veremos en breve, todos los generadores de realidad virtual lo suficientemente sofisticados tienen el mismo repertorio.

De modo que nuestro hipotético proyecto de construcción del generador de realidad virtual definitivo, que iba tan bien hasta

aquí, se encuentra de repente ante un muro infranqueable. Por más mejoras que se introduzcan en un futuro aún lejano, el repertorio de toda la tecnología de realidad virtual nunca crecerá más allá de cierto conjunto fijo de entornos. Sin duda, este conjunto es infinitamente mayor, y más diverso, en comparación con la experiencia humana anterior a la tecnología de realidad virtual, pero, no obstante, es sólo una parte infinitesimal de todos los entornos lógicamente posibles.

¿Qué se siente en un entorno cantgotu? Si bien las leyes de la física no nos permiten experimentarlo de modo directo, desde el punto de vista lógico es posible y, por tanto, legítimo preguntarse cómo sería esa experiencia. Ciertamente, no nos podría proporcionar nuevas *sensaciones*, ya que es posible un generador universal de imágenes y se asume que forma parte de nuestro generador de realidad virtual de alta tecnología. Así, un entorno cantgotu nos parecería misterioso sólo después de haberlo experimentado y reflexionar sobre los resultados. Sería algo parecido a lo siguiente: supongamos que es usted un aficionado a la realidad virtual en un lejano futuro de ultraalta tecnología. Está usted hastiado, ya que le parece que ha probado todo lo interesante en ese campo. Pero, un buen día, aparece un genio que le propone transportarle a un entorno cantgotu. Si bien al principio usted se muestra escéptico, acepta finalmente probarlo y es transportado al entorno. Tras algunos experimentos, le parece reconocerlo. Responde exactamente igual que uno de sus entornos favoritos, que en su sistema doméstico de realidad virtual tiene el número de programa X. A pesar de ello, continúa experimentando, y, en un momento dado, llegado el X.° minuto subjetivo de la experiencia, el entorno responde de un modo muy diferente a lo que se espera del entorno X. Así que abandona la idea de que se trata del entorno X. Quizás entonces se da cuenta de que todo lo sucedido es también consistente con otro entorno reproducible, el entorno Y. Pero después, durante el Y.° minuto subjetivo de la experiencia, se da cuenta de que no es así. Ésta es, simplemente, la característica de un entorno cantgotu: por más conjeturas que haga, por más complejo que sea el programa que le parezca responsable de la representación del entorno, siempre se equivocará, ya que *ningún* programa —ni de su generador de realidad virtual, ni de ningún otro— lo reproducirá.

Más pronto o más tarde, tendrá que poner fin a la prueba. Llegado a este punto, quizás se sienta inclinado a admitir que el

genio no le engañó. Esto no equivale a decir que pueda demostrar —nunca podrá— que ha estado en un entorno cantgotu, puesto que siempre habrá algún programa aún más complejo que el genio pueda ejecutar y que tenga puntos de contacto con las experiencias por las que usted ha pasado antes. Ésta es, precisamente, como ya he comentado, la característica general de la realidad virtual, es decir, que la experiencia no puede probar que uno se encuentra en un determinado entorno, sea la pista central de Wimbledon o un entorno del tipo cantgotu.

De todos modos, no existen semejantes genios ni tales entornos, así que debemos concluir que la física no permite que el repertorio de un generador de realidad virtual sea, ni con mucho, tan extenso como la sola lógica permitiría. ¿Cuán extenso puede ser?

Puesto que no podemos esperar reproducir todos los entornos lógicamente posibles, consideremos una clase más modesta (pero, en último término, más interesante) de universalidad. Definamos un *generador universal de realidad virtual* como aquel cuyo repertorio contiene el de cualquier otro generador de realidad virtual físicamente posible. ¿Puede existir semejante máquina? Sí. Ello resulta obvio —quizás demasiado obvio— sólo con pensar en los ingenios futuristas basados en la estimulación nerviosa controlada por ordenador. Dicha máquina podría ser programada para tener las características de cualquier máquina rival. Podría calcular cómo respondería la otra máquina de acuerdo con cualquier programa y cualquier comportamiento del usuario, de modo que podría reproducir estas respuestas con perfecta fidelidad (desde el punto de vista del usuario). Digo que ello resulta «quizás demasiado obvio», porque contiene una importante asunción sobre lo que el ingenio propuesto, y más específicamente su ordenador, podría ser programado para hacer: dado el programa apropiado y tiempo y espacio de almacenamiento de datos suficientes, podría calcular el resultado de cualquier cálculo realizado por cualquier otro ordenador, incluyendo el del generador de realidad virtual rival. En consecuencia, la factibilidad de un generador de realidad virtual universal depende de la existencia de un ordenador universal, una máquina única, capaz de calcular todo lo calculable.

Como he dicho, esta clase de universalidad fue estudiada al principio no por físicos, sino por matemáticos. Trataban de precisar la noción intuitiva de «calcular» (o «computar» o «demostrar»)

algo en matemáticas. No tomaron en consideración el hecho de que el cálculo matemático es un proceso físico (en particular, como he dicho, es un proceso de representación en realidad virtual) y que, por consiguiente, resulta imposible determinar mediante razonamiento matemático lo que puede ser calculado matemáticamente o no. Ello queda sometido por completo a las leyes de la física. En lugar de tratar de deducir sus resultados mediante dichas leyes, los matemáticos postularon abstractos modelos de «cálculo», y *definieron* «cálculo» y «demostración» en términos de dichos modelos (discutiré esta interesante equivocación en el capítulo 10). Así fue como en el transcurso de unos pocos meses del año 1936 tres matemáticos, Emil Post, Alonzo Church y, sobre todo, Alan Turing, crearon independientemente los primeros diseños abstractos de ordenadores universales. Cada uno de ellos conjeturaba que su modelo de «cálculo» formalizaba correctamente la noción intuitiva y tradicional de «cálculo». En consecuencia, todos ellos conjeturaban también que sus modelos respectivos equivalían a (tenían el mismo repertorio que) cualquier otra formalización razonable de la misma intuición. Esto se conoce como *conjetura de Church-Turing*.

El modelo de cálculo de Turing y su concepción de la naturaleza del problema que trataba de resolver eran los más próximos a la física. Su ordenador abstracto —*la máquina de Turing*— nacía de la idea de una cinta de papel dividida en cuadrados, en cada uno de los cuales estaba escrito un símbolo perteneciente a una serie finita de símbolos fácilmente distinguibles. El cálculo se realizaba examinando un cuadrado cada vez, haciendo avanzar o retroceder la cinta y borrando o escribiendo alguno de los símbolos según reglas sencillas e inequívocas. Turing demostró que un ordenador concreto de esta clase —*la máquina universal de Turing*— poseía el repertorio combinado de todas las demás máquinas de Turing. Conjeturó que dicho repertorio consistía, precisamente, en «cada función que fuese considerada naturalmente calculable». Se refería, por supuesto, a calculable *por matemáticos*.

Pero los matemáticos son una clase más bien atípica de objetos físicos. ¿Por qué hemos de dar por sentado que los cálculos realizados por ellos son el no va más de las tareas calculatorias? Y es que resulta que no lo son. Como explicaré en el capítulo 9, los *ordenadores cuánticos* pueden realizar cálculos que

ningún matemático (humano) será nunca, en principio, capaz de ejecutar. Está implícito en el trabajo de Turing que esperaba que lo que «fuese considerado naturalmente calculable» coincidiese con lo que fuese, al menos en principio, de naturaleza calculable. Esta esperanza equivale a una versión más amplia, física, de la conjetura de Church—Turing. El matemático Roger Penrose ha sugerido que debería ser denominada *principio de Turing*:

PRINCIPIO DE TURING (PARA ORDENADORES ABSTRACTOS QUE SIMULAN OBJETOS FÍSICOS): Existe un ordenador abstracto universal cuyo repertorio incluye todo cálculo que cualquier objeto físicamente posible pueda realizar.

Turing creía que su máquina universal era el «ordenador universal» en cuestión. Para tomar en consideración el repertorio más amplio de los ordenadores cuánticos, he enunciado el principio de modo que no especifique qué «ordenador abstracto» concreto realiza la tarea.

La prueba que he presentado de la existencia de entornos cantgotu se debe, básicamente, a Turing. Como he dicho, Turing no pensaba de manera explícita en términos de realidad virtual, pero un «entorno que pueda ser reproducido» corresponde a una clase de preguntas matemáticas cuyas respuestas pueden ser calculadas; luego, dichas respuestas son *calculables*. El resto, aquellas preguntas para las que no hay manera de calcular la repuesta, se denominan *no calculables*. Que una pregunta sea no calculable no significa que no tenga respuesta o que ésta tenga por fuerza que ser poco definida o ambigua. Bien al contrario, significa que tiene una respuesta inequívoca. Sólo que, físicamente, no hay modo, incluso en principio, de obtener dicha respuesta (o, para ser más exactos —puesto que siempre podríamos formular una conjetura afortunada pero inverificable—, de demostrar que es la respuesta). Por ejemplo, un *par primo* es una pareja de números primos cuya diferencia es dos, como 3 y 5 u 11 y l3. Los matemáticos han tratado en vano de responder a la pregunta de si existe un número finito o infinito de esas parejas. Se desconoce incluso si la pregunta es calculable. Supongamos que no. Ello equivale a decir que nadie, y ningún ordenador, podrá jamás presentar evidencias acerca de si hay un número finito o

infinito de pares primos. Aun así, la pregunta tiene una respuesta: podemos afirmar, sin duda, que, o bien existe un par primo máximo, o bien el número de pares primos es infinito. No existe una tercera posibilidad. La pregunta queda bien definida, aunque quizás no obtengamos nunca la correspondiente respuesta.

Es decir, en términos de realidad virtual: ningún generador de realidad virtual físicamente posible podrá jamás reproducir un entorno que proporcione al usuario respuestas a preguntas no computables. Un entorno así sería del tipo cantgotu. Inversamente, todo entorno cantgotu corresponde a una clase de preguntas matemáticas

(«¿Qué sucedería a continuación en un entorno definido de tal manera y tal otra»?) a las que resulta físicamente imposible responder.

Si bien las preguntas no calculables son infinitamente más numerosas que las calculables, tienden a ser más esotéricas. No es por casualidad, sino porque las partes de las matemáticas que tendemos a considerar menos esotéricas son aquellas que vemos reflejadas en el comportamiento de objetos físicos en situaciones familiares. En estos casos, a menudo podemos utilizar dichos objetos para responder a preguntas sobre las correspondientes relaciones matemáticas. Podemos, por ejemplo, contar con los dedos porque su física imita naturalmente la aritmética de los números del uno al diez.

Pronto se comprobó que los repertorios de los tres ordenadores abstractos —muy distintos entre sí— definidos por Turing, Church y Post eran idénticos. Lo mismo ha sucedido con los repertorios de todos los modelos matemáticos abstractos de cálculo que han sido propuestos desde entonces. Ello puede ser considerado un apoyo a la conjetura de Church—Turing y a la universalidad de la máquina universal de Turing. No obstante, la capacidad de cálculo de las máquinas *abstractas* no es representativa de lo que es calculable en la realidad. El ámbito de la realidad virtual, y sus más profundas implicaciones para la comprensibilidad de la naturaleza y otros aspectos de la estructura de la realidad, dependen de si los ordenadores adecuados son físicamente realizables o no. Esto nos conduce a una versión aún más amplia del principio de Turing:

PRINCIPIO DE TURING (PARA ORDENADORES FÍSICOS

QUE SE SIMULAN MUTUAMENTE): Es posible construir un ordenador universal: una máquina capaz de ser programada para realizar cualquier cálculo que cualquier otro objeto físico pueda realizar.

De ello se desprende que si un generador universal de imágenes estuviese controlado por un ordenador universal, la máquina resultante sería un generador universal de realidad virtual. En otras palabras, el siguiente principio es igualmente válido:

PRINCIPIO DE TURING (PARA GENERADORES DE REALIDAD VIRTUAL QUE SE REPRODUCEN MUTUAMENTE): Es posible construir un generador universal de realidad virtual cuyo repertorio incluya el de cualquier otro generador de realidad virtual físicamente posible.

Ahora bien, cualquier entorno puede ser reproducido por un generador de realidad virtual de *alguna* clase (por ejemplo, siempre podríamos considerar una copia de dicho entorno como un generador de realidad virtual dotado tal vez de un repertorio muy pequeño). También se desprende, pues, de esta versión del principio de Turing que cualquier entorno físicamente posible puede ser reproducido por el generador universal de realidad virtual. Por consiguiente, para expresar la muy fuerte autosemejanza que existe en la estructura de la realidad, la cual abarca no sólo los cálculos sino la totalidad de los procesos físicos, el principio de Turing puede ser enunciado del siguiente modo generalizado:

PRINCIPIO DE TURING : Es posible construir un generador de realidad virtual cuyo repertorio incluya todos los entornos físicamente posibles.

Ésta es la versión más amplia del principio de Turing. No sólo nos dice que diversas partes de la realidad pueden parecerse unas a otras, sino también que un único objeto físico, susceptible de ser construido de una vez por todas (dejando aparte el mantenimiento, así como el aporte de energía y memoria adicional cuando sea necesario), puede realizar con fidelidad ilimitada la

tarea de describir o imitar cualquier otra parte del multiverso. El conjunto de todos los comportamientos y respuestas de ese objeto refleja, con total exactitud, el conjunto de todos los comportamientos y respuestas de todos los demás objetos y procesos físicamente posibles.

Ésta es, precisamente, la clase de autosemejanza necesaria para que —de acuerdo con la esperanza manifestada en el capítulo 1— la estructura de la realidad esté verdaderamente unificada y sea comprensible. Si las leyes de la física, tal y como se aplican a objetos o procesos físicos, han de ser comprensibles, deben poder ser incorporadas a otro objeto físico: el individuo que las conoce. Es también indispensable que los procesos capaces de generar semejante conocimiento sean en sí mismos físicamente posibles. Estos procesos reciben el nombre de ciencia. La ciencia depende de la comprobación experimental, que equivale a la representación física de las predicciones de una ley y a su comparación con (una representación de) la realidad. Depende también de la explicación, lo que exige que las propias leyes abstractas —y no meramente su contenido predictivo— sean susceptibles de ser representadas en realidad virtual. Se trata de un nivel muy alto, pero la realidad lo alcanza. Mejor dicho, las leyes de la física lo alcanzan. Las leyes de la física, al conformarse al principio de Turing, hacen físicamente posible que puedan ser conocidas por objetos físicos. Así, pues, podemos decir que las leyes de la física condicionan su propia comprensibilidad.

Puesto que construir un generador universal de realidad virtual es físicamente posible, debe *estar* construido de manera real en algunos universos. Una advertencia es aquí necesaria. Como ya expliqué en el capítulo 3, podemos definir, en general, un proceso físicamente posible como aquel que ocurre de manera real en algún lugar del multiverso. Pero, hablando en sentido estricto, un generador universal de realidad virtual constituye un caso acotado dentro de unos límites, que requiere para operar recursos arbitrariamente grandes. Lo que queremos decir, pues, al considerarlo «físicamente posible» es que existen en el multiverso generadores de realidad virtual dotados de repertorios arbitrariamente próximos al conjunto de todos los entornos físicamente posibles. Del mismo modo, puesto que las leyes de la física son susceptibles de ser representadas, lo *son* en alguna parte. Se desprende así del principio de Turing (en la versión más amplia que he enunciado) que las leyes de la física no se limitan a

postular su propia comprensibilidad de algún modo abstracto —su comprensibilidad por científicos abstractos, por así decirlo—, sino que implican la existencia física, en algún lugar el multiverso, de entidades que las comprenden arbitrariamente bien. Consideraré esta implicación con más detenimiento en ulteriores capítulos.

Volvamos ahora a una cuestión planteada en el capítulo anterior: la de si, en el caso de que sólo pudiésemos aprender mediante representaciones en realidad virtual basadas en leyes físicas erróneas, serían esas leyes erróneas las que aprenderíamos. ¡Lo primero que hay que señalar es que para aprender sólo disponemos de una realidad virtual basada en leyes físicas erróneas! Como he dicho, recibimos todas nuestras experiencias externas en forma de realidad virtual generada por nuestros cerebros. Puesto que nuestros conceptos y teorías (innatos o adquiridos) nunca son perfectos, todas nuestras representaciones son por fuerza inexactas. Es decir, nos proporcionan la experiencia de un entorno significativamente distinto de aquel en que nos hallamos en realidad. Los espejismos, y otras ilusiones ópticas, son buenos ejemplos de ello. Otro ejemplo es la sensación que tenemos de que la Tierra está en reposo bajo nuestros pies, a pesar de su rápido y complejo movimiento real. Otros son que nos sentimos inmersos en un único universo y sólo vemos un aspecto de nuestro yo consciente en cada momento, cuando, en realidad, en ambos casos existen muchos. Sin embargo, estas engañosas e inexactas experiencias no proporcionan argumento alguno en contra del razonamiento científico. Bien al contrario, tales deficiencias son, precisamente, lo que lo origina.

Estamos embarcados en la resolución de problemas sobre la realidad física. Si resulta que hasta ahora hemos estudiado meramente la programación de un planetario cósmico, eso sólo significará que hemos estudiado una porción de la realidad menor de lo que creíamos. ¿Y qué? Estas cosas han ocurrido muchas veces en la historia de la ciencia a medida que nuestros horizontes se iban ampliando para incluir primero la Tierra y luego el sistema solar, la Galaxia, las otras galaxias, los cúmulos de galaxias y así sucesivamente, hasta llegar, por supuesto, a los universos paralelos. Una ampliación semejante de nuestros horizontes puede tener lugar mañana; de hecho, puede ocurrir de acuerdo con cualquiera de una infinidad de teorías, o puede no suceder nunca. En buena lógica, debemos conceder al solipsismo y a las teorías

relacionadas con él que la realidad acerca de la cual vamos aprendiendo *podría* no ser más que una fracción irrelevante de una estructura mayor, inaccesible e incomprensible. Pero la refutación general que he hecho ya de esas doctrinas demuestra que es irracional partir de esta posibilidad. De acuerdo con Occam, consideraremos esas clase de teorías cuando —y sólo cuando—, nos proporcionen mejores explicaciones que teorías rivales más sencillas.

No obstante, queda una pregunta que cabe hacerse. Supongamos que alguien se encontrara aprisionado en una pequeña e irrelevante porción de nuestra propia realidad, por ejemplo, dentro de un generador universal de realidad virtual programado con leyes de física erróneas. ¿Qué podría aprender este prisionero acerca de la realidad externa? A primera vista, parece imposible que pudiese aprender algo sobre ella. Se diría que, todo lo más, podría descubrir las leyes de funcionamiento —es decir, el programa— del ordenador que operase su prisión.

¡Pero no es así! Una vez más, debemos recordar que el prisionero, si es un científico, buscará tanto explicaciones como predicciones. En otras palabras, no se contentará, simplemente, con comprender el programa que opera su prisión, aspirará a poder explicar el origen y atributos de las diversas entidades, incluyéndose a sí mismo, que observa en la realidad que habita. Pero en la mayoría de los entornos de realidad virtual no existe tal explicación, ya que los objetos representados no se originan allí, sino que han sido diseñados en la realidad externa. Supongamos que usted utiliza un videojuego de realidad virtual. En aras de la simplicidad, supongamos que el juego es el ajedrez (una versión en primera persona, por ejemplo, en la que usted adopta la personalidad del rey). Utilizará los métodos de la ciencia para descubrir las «leyes de la física» de ese entorno, así como sus consecuencias emergentes. Aprenderá que el jaque mate y las tablas son sucesos «físicamente» posibles (es decir, posibles de acuerdo con su comprensión de cómo funciona ese entorno), pero que una posición con nueve peones blancos no es «físicamente» posible. Una vez que hubiese comprendido las leyes lo suficientemente bien, seguramente se daría cuenta de que el tablero de ajedrez es un objeto demasiado sencillo para tener pensamientos propios y, por consiguiente, los procesos de pensamiento que usted experimentase no podrían estar regidos únicamente por las leyes del ajedrez. Del mismo modo, podría

comprobar que, a lo largo de cualquier número de partidas de ajedrez, las piezas no pueden nunca evolucionar en configuraciones autorreproductoras. Y si la vida no puede evolucionar sobre el tablero, mucho menos la inteligencia, y, por lo tanto, inferiría también que sus propios procesos de pensamiento no pueden haberse originado en el universo en el que se encuentra. De este modo, y aun en el caso de haber vivido toda su vida en el interior del entorno representado y carecer, por consiguiente, de todo recuerdo del mundo exterior en el que apoyarse, su conocimiento no quedaría confinado a dicho entorno. Sabría que, aunque su universo pareciera tener cierta disposición y obedecer a determinadas leyes, debería existir un universo mucho mayor en el exterior, regido por leyes físicas distintas. Podría incluso llegar a adivinar algunas de las maneras en que dichas leyes más amplias diferirían de las del ajedrez.

Arthur C. Clarke observó que «toda tecnología lo suficientemente avanzada es indistinguible de la magia». Esto es cierto, pero un tanto engañoso. Se afirma desde la perspectiva de un pensador pre—científico, y por ello habría que invertir los términos. El hecho es que, para cualquiera que comprenda lo que es la realidad virtual, incluso la magia genuina sería indistinguible de la tecnología, ya que no hay lugar para ella en una realidad comprensible. Cualquier cosa que parezca incomprensible es vista por la ciencia, simplemente, como evidencia de que hay algo que aún no hemos comprendido, ya sea un conjuro, una tecnología avanzada o una nueva ley de la física.

El razonamiento realizado desde la premisa de la propia existencia se denomina razonamiento «antrópico». Si bien tiene cierta aplicación en cosmología, normalmente debe ser reforzado con hipótesis más consistentes acerca de la naturaleza de «uno mismo» para que pueda proporcionar conclusiones definitivas. El razonamiento antrópico no es, sin embargo, la única manera en que el preso de nuestra hipotética cárcel de realidad virtual podría adquirir conocimiento del mundo exterior. *Cualquiera* de sus explicaciones evolutivas sobre su estrecho mundo podría, en un abrir y cerrar de ojos, saltar a una realidad externa. Las propias reglas del ajedrez, por ejemplo, contienen lo que un jugador reflexivo podría reconocer como «evidencia fósil» de que dichas reglas tienen un historial evolutivo: hay movimientos «excepcionales», como el enroque y el comer al paso, que incrementan la complejidad de las reglas, pero mejoran el juego.

Al intentar explicar esa complejidad, se puede justificadamente llegar a la conclusión de que las reglas del ajedrez no han sido siempre las actuales.

En el esquema popperiano de las cosas, las explicaciones siempre conducen a nuevos problemas que, a su vez, requerirán nuevas explicaciones. Si el prisionero no consigue, pasado un tiempo, mejorar las explicaciones de que dispone, puede, por supuesto, abandonar e incluso llegar a la errónea conclusión de que no hay más explicaciones posibles. Pero si no abandona, pensará acerca de aquellos aspectos de su entorno que le parezcan inadecuadamente explicados. Por lo tanto, si los carceleros de alta tecnología quisiesen estar seguros de que su entorno representado engañase siempre a su prisionero y le indujese a creer que no existía otro mundo exterior, no acabarían nunca el trabajo. Cuanto más quisiesen prolongar la ilusión, más ingenioso debería ser el programa. No bastaría con impedir que el prisionero observara el exterior. El entorno representado debería ser tal que ninguna explicación de nada interior requiriese nunca el postulado de algo exterior. El entorno, en otras palabras, debería estar autocontenido por lo que concierne a las explicaciones. Pero dudo mucho que la realidad total, o alguna de sus partes, posea dicha propiedad.

TERMINOLOGÍA

Generador universal de realidad virtual. Aquel cuyo repertorio contiene todos los entornos físicamente posibles.

Entornos cantgotu. Entornos lógicamente posibles que no pueden ser reproducidos por ningún generador de realidad virtual físicamente posible.

Argumento diagonal. Prueba consistente en hacer una lista de un conjunto de entidades, la cual se utiliza después para demostrar la existencia de entidades relacionadas que no pueden estar incluidas en ella. Por ejemplo, una sucesión de números enteros (1, 2, 3, ...) permite demostrar la existencia de los números ¼, ½, ¾, 1¼, 1½, 1¾, etcétera, relacionados con ellos, pero que no pueden incluirse entre ellos.

Máquina de Turing. Uno de los primeros modelos abstractos de ordenador.

Máquina universal de Turing. Máquina de Turing que combina el repertorio de todas las demás máquinas de Turing.

Principio de Turing (en su versión más amplia). «Es físicamente posible construir un generador universal de realidad virtual.» De acuerdo con las hipótesis que he expuesto, esto implica que no existe límite máximo para la universalidad de los generadores de realidad virtual que puedan ser construidos realmente en algún lugar del multiverso.

SUMARIO

El argumento diagonal demuestra que una abrumadora mayoría de los entornos lógicamente posibles no puede ser reproducida mediante la realidad virtual. Les he dado el nombre de «entornos cantgotu». Existe, sin embargo, una gran autosemejanza en la realidad física de todos ellos, que queda expresada por el principio de Turing: «Es posible construir un generador de realidad virtual capaz de reproducir todos los entornos físicamente posibles.» Así pues, puede ser construido un único objeto físico capaz de imitar todos los comportamientos y respuestas de cualquier otro objeto o proceso físicamente posible. Esto es lo que hace comprensible la realidad.

Y también lo que hace posible la evolución de los organismos vivos. Sin embargo, antes de considerar la teoría de la evolución —la cuarta vía para la explicación de la estructura de la realidad—, debo hacer una breve incursión en la epistemología.

7. UNA CONVERSACIÓN SOBRE LA JUSTIFICACIÓN (O DAVID Y EL «CRIPTOINDUCTIVISTA»)

Creo haber resuelto un problema filosófico capital: el de la inducción.

Karl Popper

Como expliqué en el prefacio, este libro no constituye, básicamente, una defensa de las teorías fundamentales de las cuatro vías principales, sino una investigación sobre lo que dicen y la clase de realidad que describen. Ésta es la razón de que no me entretenga en analizar en profundidad las teorías opuestas. No obstante, hay una de esas teorías opuestas —el sentido común— que la razón me exige refutar con todo detalle cada vez que parece entrar en conflicto con aquello que afirmo. De ahí que en el capítulo 2 presentase una completa refutación de la idea de sentido común de que existe un solo universo. En el capítulo 11 haré lo mismo con la noción de que el tiempo «fluye» o nuestra conciencia «se mueve» a través de él. En el capítulo 3 critiqué el inductivismo, la idea de sentido común de que elaboramos teorías sobre el mundo físico gracias a la generalización de los resultados de las observaciones y justificamos esas teorías mediante la repetición de dichas observaciones. Expliqué que la generalización induc—tivista a partir de la observación es imposible, y que su justificación no es válida. Demostré que el inductivismo se apoya en la idea equivocada de que la ciencia busca predicciones basadas en observaciones, cuando, en realidad, busca explicaciones como respuesta a problemas. Expuse también (siguiendo a Popper) cómo progresa la ciencia: conjeturando nuevas explicaciones y seleccionando luego las mejores para experimentarlas. Todo ello ha sido mayoritariamente aceptado por científicos y filósofos de la ciencia. Lo que, en cambio, no lo ha sido, es que este proceso se halle *justificado*. Me explicaré.

La ciencia busca mejores explicaciones. Una explicación científica ilustra nuestras observaciones postulando algo sobre qué

es la realidad y cómo funciona. Consideramos que una explicación es mejor que otra cuando deja menos cabos sueltos (tales como entidades cuyas propiedades no son explicadas), requiere menos postulados y éstos son más sencillos, es más general, se combina con facilidad con buenas explicaciones en otros campos, etcétera. Pero ¿por qué razón ha de ser una explicación mejor lo que en la práctica damos por sentado, es decir, la señal de que una teoría es *más cierta*? ¿Y por qué, inviniendo los términos, ha de ser por fuerza falsa una explicación indudablemente mala (por ejemplo, una que no tuviera ninguno de los atributos mencionados antes)? En efecto, desde un punto de vista lógico no tiene por qué haber conexiones entre la verdad y la capacidad explicativa. Una mala explicación (como el solipsismo) *puede* ser cierta. E incluso la mejor y más fiable de las teorías puede dar una predicción equivocada en algunas ocasiones, quizás aquellas en que más confiemos en ella. Ninguna forma válida de razonamiento puede no ya desechar de manera lógica semejantes posibilidades, sino ni siquiera demostrar que son improbables. Pero, si es así, ¿qué justifica nuestra confianza en las mejores explicaciones como guías para la toma práctica de decisiones? Y, de modo más general, cualesquiera que sean los criterios que utilicemos para juzgar a las teorías científicas, ¿cómo es posible que el hecho de que una teoría satisfaga dichos criterios en la actualidad nos induzca a creer que seguirá satisfaciéndolos en un futuro más o menos lejano?

Ésta es la versión moderna del «problema de la inducción». La mayoría de los filósofos están de acuerdo en la actualidad con el argumento de Popper de que las nuevas teorías no se infieren de nada, sino que son meras hipótesis. Aceptan también que el progreso científico se realiza mediante conjeturas y refutaciones (como describí en el capítulo 3), y que las teorías son aceptadas cuando sus rivales son refutadas y no en virtud de numerosas observaciones que las confirmen. Aceptan que el conocimiento así obtenido suele tender a ser fiable. El problema es que no ven el porqué. Los inductivistas tradicionales trataron de formular un «principio de inducción» que afirmaba que las observaciones que la confirmaban hacían a una teoría más plausible, o que «el futuro se parecerá al pasado», o cualquier enunciado por el estilo. Trataron también de formular una metodología científica inductiva desarrollando leyes que especificaban qué clase de inferencias se podían extraer válidamente de los «datos». Fracasaron en todo,

por las razones que ya he explicado. Pero, aun en el caso de que lo hubiesen conseguido, en el sentido de ser capaces de elaborar un esquema que pudiese ser seguido con éxito para crear conocimiento científico, ello no habría resuelto el problema de la inducción tal como se plantea hoy día, ya que, en este caso, la «inducción» sería, simplemente, una de las posibles maneras de escoger teorías, y el problema seguiría siendo *por qué estas teorías han de constituir bases fiables para la acción*. En otras palabras, los filósofos que se preocupan actualmente por el «problema de la inducción» no son inductivistas a la antigua usanza. No tratan de obtener o justificar ninguna teoría inductivamente. No creen que se hunda el cielo, pero no saben cómo justificar esta confianza.

Los filósofos suspiran hoy por esa justificación perdida. No creen ya que la inducción pueda proporcionarla, pero, sin embargo, sienten un vacío en forma de inducciónismo en su concepción del mundo, al igual que las personas religiosas que han perdido la fe sufren de un «vacío en forma de Dios» en *su* concepción del mundo. Pero, en mi opinión, hay poca diferencia entre sentir un vacío en forma de X en la personal concepción del mundo y creer en X. Por ello, para que encaje mejor en el marco de este concepto más complejo del problema de la inducción, deseo redefinir el término «inductivista» y utilizarlo para designar a alguien que piensa que la *invalidez* de la justificación inductiva constituye un problema para los fundamentos de la ciencia. Dicho de otro modo, un inductivista cree que hay un vacío, y que debe ser llenado, por el principio de inducción o por lo que sea. A algunos inductivistas no les importa que los tachen de tales. Pero a otros sí; son los que denominaré *criptoinductivistas*.

La mayor parte de los filósofos contemporáneos son criptoinductivistas. Y lo que empeora aún más las cosas es que (al igual que la mayoría de los científicos) subestiman de un modo inexcusable el papel de la explicación en el proceso científico. Lo mismo hace la mayoría de los antiinductivistas popperianos, inclinados a negar que exista la justificación ni nada que se le parezca (ni siquiera a título experimental), lo cual abre un nuevo vacío explicativo en su concepción del mundo. El filósofo John Worrall ha expuesto el problema, tal como él lo ve, en un diálogo imaginario entre Popper y otros filósofos titulado «Why Both Popper and Watkins Fail to Solve the Problem of Induction» [Por qué ni Popper ni Watkins consiguen resolver el problema de la

inducción][3]. La escena se desarrolla en la cúspide de la Torre Eiffel. Uno de los participantes —el Planeador—decide saltar al vacío en vez de utilizar el ascensor. Los demás intentan persuadirlo de que saltar significa una muerte segura. Utilizan para ello los mejores argumentos científicos y filosóficos de que disponen. Pero el exasperante Planeador sigue confiando en que descenderá planeando suavemente, sin sufrir el menor daño, y argumenta que, de acuerdo con las experiencias anteriores, no puede probarse lógicamente que una expectativa contraria a la suya pueda ser preferible.

Creo que podemos justificar nuestra expectativa de que el Planeador se mataría. La justificación —siempre a título experimental, desde luego— proviene de las explicaciones proporcionadas por las teorías de eminentes científicos. Hasta el punto en que estas teorías son buenas, está racionalmente justificado fiarse de sus correspondientes predicciones. Así pues, y como respuesta a Worrall, presento aquí mi propio diálogo, que se desarrolla en el mismo lugar:

DAVID: Desde que leí lo que dice Popper sobre la inducción, creo que realmente solucionó, como proclama, el problema de la inducción. No obstante, pocos filósofos están de acuerdo. ¿Por qué?

CRIPTOINDUCTIVISTA: Porque Popper nunca se planteó el problema de la inducción tal como nosotros lo entendemos. Lo que hizo fue presentar una crítica del *inductivismo*. El inductivismo afirmaba que existe un modo «inductivo» de razonamiento que, partiendo de la evidencia de observaciones efectuadas en el pasado, permite enunciar teorías generales sobre el futuro y justificar su uso. Mantenía que hay un principio natural, el *principio de inducción*, que dice, más o menos, que «las observaciones hechas en el futuro se parecerán a las realizadas en el pasado en circunstancias similares». Hubo varios intentos de formularlo de tal manera que permitiese, en efecto, enunciar teorías o justificarlas a partir de observaciones individuales. Todos fracasaron. La crítica de Popper, si bien influyente entre los científicos, sobre todo, en combinación con sus trabajos de elucidación de la metodología de la ciencia, no era original, ni mucho menos. La endeblez del inductivismo era conocida desde sus orígenes y, por supuesto, desde las críticas de David Hume a principios del siglo XVIII. El

problema de la inducción no consiste en cómo justificar o refutar el principio de inducción, sino, más bien —dando por sentado que no es válido—, en *cómo justificar cualquier conclusión sobre el futuro a partir de observaciones realizadas en el pasado*. Y antes de que me diga que no es necesario...

DAVID: No es necesario.

CRIPTOINDUCTIVISTA: Sí que lo es. Esto es lo que resulta más irritante de ustedes, los popperianos: que nieguen lo evidente. Obviamente, la razón de que no salte usted por encima de esta barandilla es, en parte, por que considera *justificado* confiar en la mejor teoría de la gravedad e *injustificado* confiar en algunas otras. Por supuesto, con «la mejor teoría de la gravedad» quiero decir, en este caso, algo más que la simple teoría de la relatividad general. Me refiero también al complejo conjunto de teorías sobre fenómenos tales como resistencia del aire, fisiología humana, elasticidad del hormigón y disponibilidad de medios de rescate aéreos a baja altura.

DAVID: Sí, consideraría justificado confiar en esa teoría. De acuerdo con la metodología popperiana, en casos así deberíamos confiar en la teoría *mejor corroborada*, es decir, la que ha superado las pruebas más estrictas, que sus rivales no han podido pasar.

CRIPTOINDUCTIVISTA: Dice que «*deberíamos* confiar en la teoría mejor corroborada», pero ¿por qué, exactamente? Puedo presumir que porque, según Popper, el proceso de corroboración la ha justificado, en el sentido de que sus predicciones tienen más probabilidades de ser ciertas que las de otras teorías.

DAVID: Bueno, no es que tengan más probabilidades de ser ciertas que las de *todas* las demás teorías, ya que, sin duda, algún día dispondremos de mejores teorías de la gravedad....

CRIPTOINDUCTIVISTA: ¡Un momento! Pongámonos de acuerdo, por favor, en no zancadillearnos mutuamente con observaciones de poca monta, sin importancia para la sustancia de lo que estamos discutiendo. *Por supuesto* que algún día habrá una teoría mejor de la gravedad, pero usted debe decidir si salta ahora. Dada la evidencia de que disponemos actualmente, ha escogido obrar según una determinada teoría, y lo ha hecho de acuerdo con los criterios popperianos, puesto que cree que son los más adecuados para seleccionar teorías que ofrezcan predicciones ciertas.

DAVID: Sí.

CRIPTOINDUCTIVISTA: Así que, resumiendo, cree que la evidencia de que disponemos actualmente justifica la predicción de que se mataría en caso de saltar por encima de la barandilla.

DAVID: No, no lo hace.

CRIPTOINDUCT1VISTA: ¡Pero hombre, se está contradiciendo! Acaba de afirmar que esa predicción *está* justificada.

DAVID: Y lo está, pero no por la evidencia, si por «evidencia» entiende usted todos los experimentos cuyos resultados predijo acertadamente la teoría en el pasado. Como todos sabemos, esa evidencia es consistente con infinidad de teorías, incluyendo aquellas que predicen todos los desenlaces lógicamente posibles de mi salto por encima de la barandilla.

CRIPTOINDUCTIVISTA: Así pues, en vista de lo que acaba de decir, repito que todo el problema consiste en encontrar qué es lo que justifica la predicción. Ése es el problema de la inducción.

DAVID: Bueno, ése es el problema que Popper resolvió.

CRIPTOINDUCTIVISTA: Esto es nuevo para mí, y, créame, he estudiado a Popper a fondo. Pero, en todo caso, ¿cuál es la solución? Estoy ansioso por escucharla. ¿Qué justifica la predicción, si no es la evidencia?

DAVID: La argumentación.

CRIPTOINDUCTIVISTA: ¿La argumentación?

DAVID: Sólo la argumentación puede justificar algo, aunque siempre a título experimental, por supuesto. Toda teorización está sujeta a error, etcétera, etcétera. Pero, aun así, la argumentación puede, en ocasiones, justificar teorías. Es su razón de ser.

CRIPTOINDUCTIVISTA: Creo que ésta es otra de sus sutilezas. No puede decir en serio que la teoría puede quedar justificada mediante la *pura* argumentación, como un teorema matemático[4]. Sin duda, la evidencia tiene también algún papel.

DAVID: Sin duda. Se trata de una teoría empírica, y, por consiguiente, según la metodología científica popperiana, los experimentos cruciales tienen un papel fundamental para decidir entre ella y sus rivales. Éstas fueron refutadas y aquélla sobrevivió.

CRIPTOINDUCTIVISTA: Y, como consecuencia de la refutación de las unas y la supervivencia de la otra, todo lo cual sucedió en el pasado, por cierto, la utilización práctica de esa teoría para predecir el futuro está ahora justificada.

DAVID: Supongo que sí, aunque me parece capcioso decir

«como consecuencia de» cuando no estamos hablando de una deducción lógica.

CRIPTOINDUCTIVISTA: Bueno, eso nos lleva a otra pregunta importante: *¿Qué clase de consecuencia era?* ¡A ver si le pongo en un aprieto! Admite que tanto la argumentación *como* los resultados de los experimentos justificaron la teoría. Si los experimentos hubiesen tenido resultados distintos, la argumentación habría justificado una teoría diferente. ¿Acepta, pues, que en este sentido, es decir, por la vía de la argumentación, los resultados de los experimentos pasados justifican la predicción?

DAVID: Sí.

CRIPTOINDUCTIVISTA: Entonces, ¿qué era lo que tenían esos resultados pasados reales que justificaron la predicción, la comparación con otros resultados pasados posibles, que hubieran podido justificar la predicción contraria?

DAVID: Lo que tenían los resultados reales era que refutaron todas las teorías rivales y corroboraron la que ahora prevalece.

CRIPTOINDUCTIVISTA: Muy bien. Ahora escúcheme con atención, porque acaba de decir algo que no sólo es demostrablemente falso, sino que usted mismo lo reconoció así hace unos momentos. Ha dicho que los resultados de los experimentos «refutaron todas las teorías rivales». Pero, como muy bien sabe, ningún conjunto de resultados de experimentos puede refutar todas las posibles teorías rivales de una teoría general. Usted mismo ha dicho que cualquier conjunto de resultados pasados es, según sus palabras textuales, «consistente con infinidad de teorías, incluyendo aquellas que predicen todos los desenlaces lógicamente posibles de mi salto por encima de la barandilla». De ello se desprende, inexorablemente, que la predicción por la que se inclina *no estaba* justificada por los resultados experimentales, puesto que hay infinidad de teorías rivales de la suya, asimismo todavía sin refutar, que hacen la predicción contraria.

DAVID: Me alegro de haberle escuchado con atención, como me pidió, porque así he podido comprender que buena parte de esta discusión se debe a que utilizamos una terminología distinta. Cuando Popper habla de «rivales» de una determinada teoría, no se refiere al conjunto de todas las teorías lógicamente posibles que compiten con ella, sino sólo a las verdaderas rivales, propuestas en el curso de una controversia racional. Esto incluye las teorías

«propuestas» de modo puramente mental por una persona en el curso de «controversias» mantenidas en su cerebro.

CRIPTOINDUCTIVISTA: Ya entiendo. De acuerdo, aceptaré su terminología. Pero, incidentalmente, pues, aunque no creo que sea importante para lo que nos ocupa ahora, me inspira curiosidad, ¿no resulta una extraña aserción esa que atribuye a Popper de que la fiabilidad de una teoría depende de algo tan accidental como qué teorías, falsas teorías, han sido propuestas por otros científicos en el pasado, y no exclusivamente de su contenido y de la evidencia experimental?

DAVID: No lo creo. Incluso ustedes, los inductivistas, hablan de...

CRIPTOINDUCTIVISTA: ¡No soy inductivista!

DAVID: Sí que lo es.

CRIPTOINDUCTIVISTA: Muy bien. Aceptaré una vez más su terminología, si insiste, pero igual podría llamarme puercoespín. Es realmente inicuo llamar «inductivista» a alguien porque sostiene la tesis de que la *invalidez* del razonamiento inductivo nos enfrenta a un problema filosófico sin solucionar.

DAVID: Opino lo contrario. Pienso que esa tesis es la que define, y siempre ha definido, a un inductivista. Pero veo que Popper ha conseguido, al menos, una cosa: ¡«inductivista» se ha convertido en insulto! Pero sigamos. Le estaba explicando por qué no resulta sorprendente que la fiabilidad de una teoría dependa de qué falsas teorías hayan sido propuestas con anterioridad. Incluso los inductivistas dicen que una teoría es fiable o no de acuerdo con cierta «evidencia». Pues bien, los popperianos podrían decir que una teoría es la mejor de que se dispone para utilizarla en la práctica de acuerdo con cierto *problema-situación*. Y las características más relevantes de un problema-situación son: qué teorías y explicaciones compiten, qué argumentación ha sido presentada y qué teorías han sido refutadas. La «corroboración» no es tan sólo la confirmación de la teoría ganadora. Requiere la refutación experimental de las teorías rivales. Los hechos que confirman algo, por sí mismos, son irrelevantes.

CR1PTOINDUCTIV1STA: Muy interesante. Ahora comprendo el papel de las teorías rivales refutadas en la justificación de las predicciones de la teoría seleccionada. Según el inductivismo, se daba prioridad a la observación. Se tomaba un conjunto de observaciones pasadas, del cual se suponía que se deducía la teoría, y las observaciones constituían también la evidencia que,

de algún modo, justificaba dicha teoría. En la concepción popperiana del progreso de la ciencia, lo primordial no son las observaciones, sino los problemas, la controversia, las teorías y la crítica. Los experimentos se idean y realizan únicamente para resolver controversias. Por consiguiente, sólo los resultados experimentales que realmente refuten una teoría, y no una teoría cualquiera, sino una teoría que haya sido competidora válida en una controversia racional, pueden considerarse «corroboración». Así pues, son tan sólo esos experimentos los que arrojan evidencia en favor de la fiabilidad de la teoría vencedora.

DAVID: Correcto. Pero, incluso entonces, la «fiabilidad» que confiere la corroboración no es absoluta, sino sólo en relación con las demás teorías contendientes. Es decir, esperamos que la estrategia de confiar en teorías corroboradas nos ayude a seleccionar las mejores teorías entre las propuestas. Esto ya es suficiente para proceder. No necesitamos, ni podríamos obtener válidamente, ninguna seguridad acerca de lo *buena* que pueda ser una línea de acción propuesta, incluso la mejor. Es más, podemos equivocarnos siempre, pero ¿y qué? No podemos utilizar teorías que aún no han sido propuestas, ni enmendar errores que aún no hemos cometido.

CRIPTOINDUCTIVISTA: Cierto. Me alegro de haber aprendido algo sobre metodología científica. Pero ahora, y espero que no lo considere una descortesía, no puedo menos que volver a plantear la pregunta que he estado haciendo todo el tiempo. Supongamos que una teoría ha pasado por todo ese proceso. De vez en cuando tuvo rivales, se realizaron experimentos y todas sus rivales fueron refutadas. Como ella no lo fue, quedó corroborada. *¿Qué es lo que justifica que, por el hecho de estar corroborada, confiemos en ella en elfuturo?*

DAVID: Puesto que todas sus rivales han sido refutadas, ya no son racionalmente sostenibles. La teoría corroborada es la única que queda que puede sostenerse desde un punto de vista racional.

CRIPTOINDUCTIVISTA: Pero eso no es más que cambiar de enfoque y concentrarse en el significado futuro de la corroboración pasada en vez de hacerlo en el significado futuro de la refutación pasada. El problema sigue sin resolver. ¿Por qué, exactamente, es «insostenible desde un punto de vista racional» una teoría refutada por la experimentación? ¿Es que el hecho de que tenga una sola consecuencia falsa implica que no puede ser cierta?

DAVID: Sí.

CRIPTOINDUCTIVISTA: Pero, sin duda, por lo que se refiere a la futura aplicabilidad de la teoría, esta crítica no es relevante. Por supuesto, una teoría refutada no podrá ser *universalmente*[5] cierta. Es evidente que no puede haber sido cierta en el pasado, cuando fue experimentada, pero podría, sin embargo, tener todavía muchas consecuencias ciertas y, por qué no, ser universalmente cierta en el futuro.

DAVID: Esta terminología de «cierta en el pasado» y «cierta en el futuro» es engañosa. Cada predicción específica de una teoría es cierta o es falsa, y eso no puede cambiar. Lo que usted quiere decir, en realidad, es que, si bien la teoría refutada es estrictamente falsa, puesto que hizo algunas falsas predicciones, todas sus predicciones sobre el futuro podrían, sin embargo, resultar ciertas. En otras palabras, podría ser cierta una *teoría diferente*, que hiciera las mismas predicciones para el futuro, pero que hubiera hecho unas predicciones distintas en el pasado.

CRIPTOINDUCTIVISTA: Como quiera. Así, en vez de preguntar por qué una teoría refutada es insostenible desde un punto de vista racional, hablando con propiedad debería preguntar lo siguiente: ¿por qué la refutación de una teoría también convierte en insostenible a cualquiera de sus variantes que esté de acuerdo con ella sobre el futuro, incluso si la variante no ha sido refutada?

DAVID: No es que la refutación *convierta* en insostenibles a estas teorías; en algunas ocasiones ya *son* insostenibles por el hecho de ser malas explicaciones. Y es en estos casos cuando la ciencia puede progresar. Para que una teoría venza en una confrontación, todas sus rivales deben ser insostenibles, y ello incluye todas las variantes de las rivales *en las que alguien haya pensado*. Pero recordemos que son tan sólo las rivales en las que alguien haya pensado las que serán insostenibles. En el caso de la gravedad, por ejemplo, nadie ha propuesto nunca una teoría sostenible que esté de acuerdo con la dominante en todas sus predicciones comprobadas, pero difiera de ella en sus predicciones sobre futuros experimentos. Estoy seguro de que tales teorías son posibles. Sin ir más lejos, la sucesora de la teoría dominante será una de ellas. Pero si nadie ha pensado aún en una teoría así, ¿cómo podemos tomarla en consideración?

CRIPTOINDUCTIVISTA: ¿Qué quiere decir con eso de que «nadie ha pensado aún en una teoría así»? Yo podría pensar en

una ahora mismo, sin la menor dificultad.

DAVID: Lo dudo.

CRIPTOINDUCTIVISTA: ¡Por supuesto que puedo! Ahí va. «Siempre que usted, David, salte desde una altura, lo que, de acuerdo con la teoría dominante, debería matarlo, en vez de estrellarse, planeará. Por lo demás, la teoría en vigor se mantiene universalmente.» No puede negarme que todas las comprobaciones pasadas de su teoría lo fueron por fuerza de la mía, puesto que todas las predicciones de ambas para experimentos pasados coinciden. Así pues, las rivales refutadas de su teoría lo son también de la mía. Por consiguiente, mi nueva teoría ha sido corroborada exactamente del mismo modo que su teoría dominante. ¿Cómo es posible, entonces, que la mía sea «insostenible»? ¿Qué defectos tiene que no comparta la suya?

DAVID: ¡Precisamente, todos los mencionados en el libro de Popper! Su teoría está enunciada añadiendo un apéndice, en forma de la restricción inexplicada de que planearé, a una teoría en vigor. Esa restricción es, en efecto, una nueva teoría, pero no ha presentado usted ninguna argumentación contra la teoría dominante acerca de mis propiedades gravitatorias ni en favor de la nueva teoría. No la ha sometido a ninguna crítica, si se exceptúa la que estoy haciendo en este momento, ni a experimentación. No soluciona, ni se lo propone, ningún problema vigente, ni plantea ningún problema nuevo e interesante que pudiese solucionar. Y, lo que es peor, su restricción no sólo no explica nada, sino que *desbarata* la explicación sobre la gravedad que constituye la base de la teoría dominante. Es esa explicación la que justifica que confiemos en dicha teoría y no en la suya. Luego, de acuerdo con todos los criterios racionales, la restricción que propone puede ser rechazada de plano.

CRIPTOINDUCTIVISTA: ¿Y yo no podría decir exactamente lo mismo sobre su teoría? Sólo difiere de la mía por la misma restricción menor, pero formulada a la inversa. Dice que debería haber explicado mi restricción, pero ¿por qué no han de ser simétricas nuestras posiciones?

DAVID: Porque su teoría no va acompañada de una explicación sobre sus predicciones, mientras que la mía sí.

CRIPTOINDUCTIVISTA: Pero si mi teoría hubiese sido propuesta antes que la suya, habría sido ésta la que habría parecido contener una restricción inexplicada, y, por consiguiente, habría sido «rechazada de plano».

DAVID: Eso es, simplemente, una falacia. Cualquier persona racional que comparase su teoría con la dominante, aunque la suya hubiese sido formulada con anterioridad, la rechazaría inmediatamente en beneficio de la establecida, puesto que resulta evidente, por su mismo planteamiento, el hecho de que su teoría es una modificación inexplicada de otra.

CRIPTOINDUCTIVISTA: ¿Quiere decir que mi teoría reviste la forma de «dicha teoría es universalmente válida, excepto en tal o cual situación», pero no explico por qué rige esa excepción?

DAVID: Exactamente.

CRIPTOINDUCTIVISTA: ¡Ajá! Bien, creo que puedo probar, con la ayuda del filósofo Nelson Goodman, que está equivocado en eso. Consideremos una variante del castellano que no tenga el verbo «caer». En su lugar tiene el verbo «x-caer», que significa caer, excepto cuando se refiere a usted: entonces significa «planear». De manera semejante, «x-planear» significa «planear», excepto cuando se refiere a usted: entonces significa «caer». En este nuevo lenguaje, podría expresar mi teoría mediante la afirmación no restrictiva «todos los objetos x-caen si no son sostenidos». Pero la teoría dominante, que dice en castellano normal que «todos los objetos caen si no son sostenidos», diría en el nuevo lenguaje que «todos los objetos x-caen si no son sostenidos *excepto David, que x-planea*». Así pues, cuál de las dos teorías se considere restrictiva dependerá del lenguaje en que esté expresada ¿no es así?

DAVID: Formalmente, sí. Pero eso es intrascendente. Su teoría contiene, *en sustancia*, una afirmación inexplicada que restringe la teoría dominante. Ésta es, *en sustancia*, su misma teoría, despojada de una restricción inexplicada. Lo mire por donde lo mire, es un hecho objetivo, independiente del lenguaje.

CRIPTOINDUCTIV1STA: No veo por qué. Usted mismo ha empleado la *forma* de mi teoría para señalar la «innecesaria restricción». Dijo que resultaba «evidente» como una cláusula adicional en mi propia formulación de la teoría... en castellano. Pero cuando ésta es traducida al nuevo lenguaje, no se manifiesta restricción alguna, y, en cambio, aparece una restricción manifiesta en la formulación de la teoría dominante.

DAVID: Sí, en efecto, pero no todos los lenguajes son iguales. *Los lenguajes son teorías*. En su vocabulario y su gramática están implícitas importantes aserciones sobre el mundo. Cuando formulamos una teoría, sólo una pequeña parte de su

contenido es explícito. El resto es transmitido por el lenguaje. Como toda teoría, los lenguajes se inventan y seleccionan por su capacidad para resolver determinados problemas. En este caso, los problemas son los de expresar otras teorías en formas convenientes para su aplicación, y compararlas y criticarlas. Una de las principales maneras en que los lenguajes solucionan esos problemas es asumiendo implícitamente otras teorías incontrovertidas y que se dan por sentadas, y permitiendo al mismo tiempo que se exprese explícita y claramente todo aquello que necesite ser planteado o discutido.

CRIPTOINDUCTIVISTA: Lo acepto.

DAVID: No es casual, pues, que un lenguaje opte por una serie de nociones, y no por otra, para cubrir un determinado marco conceptual. Refleja así el estado actual en que se encuentra el problema-situación de quien habla. Por ello, la forma de su teoría, *en castellano*, es un buen indicador de su status respecto del *actual* problema-situación, tanto si soluciona problemas como si los exacerba. Pero no me quejo de la forma de su teoría, sino de su sustancia. Mi recriminación consiste en que su teoría no resuelve nada y sólo exacerba el problema-situación. Este defecto es evidente cuando la teoría es formulada en castellano, y está implícito cuando lo es en su nuevo lenguaje, pero no por ello deja de ser menos grave. Podría manifestar mi crítica igualmente bien en castellano que en lenguaje científico, en su nuevo lenguaje particular o en cualquier otro capaz de expresar la discusión que estamos manteniendo. Uno de los axiomas popperianos es que siempre debemos estar dispuestos a desarrollar la discusión en la terminología de nuestro oponente.

CRIPTOINDUCTIVISTA: Puede que tenga razón en eso. Pero ¿podría explicarlo un poco más? ¿De qué modo exacerba mi teoría el problema-situación, y por qué resultaría ello evidente incluso para alguien que tuviera mi hipotético lenguaje como propio?

DAVID: Su teoría afirma la existencia de una *anomalía* física, inexistente según la teoría dominante. Dicha anomalía sería mi supuesta inmunidad a la gravedad. Ciertamente, puede usted inventarse un lenguaje que exprese de manera implícita esa anomalía, de modo que los enunciados de su nueva teoría no tengan que referirse de modo explícito a ella. Pero, sea como fuere, seguirán refiriéndose a ella. Una rosa, aun con otro nombre, olerá igual de bien. Supongamos que usted —o cualquier otra persona— fuese hablante de su lenguaje y creyese que su teoría

de la gravedad es cierta. Supongamos que todos la diéramos por sentada, y fuese para nosotros tan natural, que usáramos la palabra «x-caer» para describir lo que usted o yo haríamos en caso de saltar por encima de la barandilla. Nada de ello altera en lo más mínimo la obvia diferencia que existiría entre mi respuesta a la gravedad y la de todos los demás. Seguro que me envidiaría si tuviese la desgracia de asomarse a esta barandilla y caerse al vacío. Lo más probable es que pensase: «¡Si pudiese responder a la gravedad como David, en vez del modo completamente distinto en que lo estoy haciendo!»

CRIPTOINDUCTIVISTA: Cierto. Sólo porque la misma palabra, «x-caer», describe su respuesta y la mía a la gravedad, no pensaría que la respuesta real sería la misma. Bien al contrario, al ser un hablante de mi hipotético lenguaje, sabría muy bien que «x-caer» es físicamente distinto para usted y para mí, del mismo modo que alguien que hable castellano sabe muy bien que no es físicamente lo mismo decir «está agotado» refiriéndose a una persona o a un pozo de petróleo. No pensaría: «Si esto le hubiese pasado a David, estaría x-cayendo como yo», sino «Si esto le hubiese pasado a David, x-caería y sobreviviría, mientras que yo estoy x-cayendo y moriré.»

DAVID: Y, lo que es más, a pesar de estar seguro de que yo planearía, *no entendería usted por qué.* Saber no es lo mismo que entender. Sentiría curiosidad por entender la explicación de esta «bien conocida» anomalía, y, como usted, todo el mundo. Los físicos acudirían, procedentes de los más apartados rincones de la Tierra, para estudiar mis anómalas propiedades gravitatorias. De hecho, si su hipotético lenguaje fuese realmente el establecido y su teoría estuviese admitida de modo general, ¡el mundo científico habría estado esperando con impaciencia mi nacimiento, y haría cola para tener el privilegio de dejarme caer desde un aeroplano! Pero, por supuesto, la premisa de todo esto, es decir, que su teoría estuviese admitida e incorporada al lenguaje establecido, es absurda. Con teoría o sin ella, con lenguaje o sin él, en la vida real ninguna persona racional consideraría siquiera la posibilidad de semejante anomalía física sin que existiese una explicación muy poderosa en su favor. Por lo tanto, su teoría sería rechazada de pleno, al igual que su hipotético lenguaje, que no es más que otro modo de formularla.

CRIPTOINDUCTIVISTA: ¿No podría ser que hubiese aquí, después de todo, una solución oculta al problema de la inducción?

Veamos. ¿En qué cambia las cosas esa consideración intuitiva de la falsedad de mi lenguaje? Mi argumentación se apoyaba en una aparente simetría entre su posición y la mía. Ambos hemos adoptado teorías coherentes con resultados experimentales existentes y cuyas rivales, excepto la una respecto de la otra, han sido refutadas. Dijo que me mostraba irracional porque mi teoría incorporaba una afirmación inexplicada, a lo que argumenté que, en un lenguaje distinto, sería su teoría la que contendría la afirmación inexplicada, de modo que la simetría seguía existiendo. Me dice ahora que los lenguajes son teorías, y que la combinación de mi teoría con mi hipotético lenguaje afirma la existencia de una anomalía física, objetiva, en comparación con la afirmación hecha por la combinación del castellano y la teoría dominante. Es aquí donde tanto la simetría entre nuestras posiciones como la argumentación que he expuesto se hunden sin remedio.

DAVID: Sin duda alguna.

CRIPTOINDUCTIVISTA: Veamos si soy capaz de aclarar las cosas un poco. ¿Me está diciendo que constituye un principio de racionalidad el hecho de que una teoría que afirma la existencia de una anomalía física, objetiva, es, si por lo demás son iguales, menos susceptible de hacer predicciones acertadas que otra que no lo afirma?

DAVID: No exactamente. Es menos probable que hagan predicciones acertadas las teorías que postulan anomalías *sin explicarlas* que sus rivales. Más generalmente, es un principio de racionalidad que las teorías se postulen para resolver problemas. Por lo tanto, *cualquier* postulado que no resuelva un problema, debe ser rechazado. Por esta razón, una buena explicación, restringida por un postulado así, se convierte en una mala explicación.

CRIPTOINDUCTIVISTA: Ahora que ya comprendo que hay realmente una diferencia objetiva entre las teorías que formulan predicciones inexplicadas y las que no, debo admitir que ello parece prometedor para la solución del problema de la inducción. Se diría que ha descubierto un modo de justificar su futura confianza en la teoría de la gravedad basado únicamente en el problema—situación pasado, lo que incluye la evidencia proporcionada por las observaciones pasadas, y la distinción entre una buena y una mala explicación. No necesita hacer ninguna afirmación como «el futuro se parecerá, probablemente, al pasado».

DAVID: No fui yo quien lo descubrió.

CRIPTOINDUCTIVISTA: Bueno, tampoco creo que fuera Popper. La razón es que no creía que las teorías científicas pudiesen ser *justificadas* de ningún modo. Usted establece una cuidadosa distinción entre las teorías que son justificadas por observaciones (como piensan los inductivistas) y las que lo son por la argumentación, pero Popper no la establecía, y, en relación con el problema de la inducción, dijo textualmente que, si bien las futuras predicciones de una teoría no pueden ser justificadas, ¡deberíamos obrar como si lo fuesen!

DAVID: No creo que fuera eso lo que dijo exactamente. Y, si lo dijo, sin duda, no es lo que quería decir.

CRIPTOINDUCTIVISTA: ¿Qué?

DAVID: Y, si es lo que quería decir, se equivocaba. ¿Por qué se sorprende tanto? Es perfectamente posible que una persona descubra una nueva teoría, en este caso la epistemología pop—periana, y, sin embargo, conserve creencias que la contradicen. Cuanto más profunda sea la teoría, más probable es que eso suceda.

CRIPTOINDUCTIVISTA: ¿Pretende entender la teoría de Popper mejor que él?

DAVID: Ni lo sé, ni me importa. La reverencia que muestran los filósofos por las fuentes históricas de las ideas me parece absurda, ¿sabe? En ciencia no consideramos que el descubridor de una teoría sea quien mejor pueda explicarla. Por ello, rara vez consultamos las fuentes originales. Invariablemente, se vuelven obsoletas a medida que los problemas—situaciones que las desencadenaron se transforman por obra de los propios descubrimientos. Por poner un ejemplo, la mayoría de los teóricos de la relatividad actuales comprenden mejor la teoría de Einstein que él cuando la descubrió. Los fundadores de la teoría cuántica se hicieron un buen lío intentando entenderla. Inicios así de vacilantes son de esperar. Si nos encaramamos sobre los hombros de gigantes, no es de extrañar que veamos más lejos que ellos. Pero, en cualquier caso, es más interesante, sin duda, argumentar acerca de dónde está la verdad que acerca de lo que creía o no un determinado pensador, por grande que éste sea.

CRIPTOINDUCTIVISTA: Muy bien, estoy de acuerdo. Pero espere un momento. Creo que me precipité cuando dije que usted no postulaba ninguna clase de principio de inducción. Verá: ha justificado que una teoría sobre el futuro, la teoría dominante de la

gravedad, es más fiable que otra, la que yo he propuesto, no obstante ser ambas coherentes con las observaciones actualmente conocidas. Puesto que la teoría dominante es de aplicación tanto para el futuro como para el pasado, usted ha justificado la proposición de que, en lo concerniente a la gravedad, *el futuro se parecerá al pasado*. Lo mismo sucederá cada vez que justifique una teoría como fiable basándose en que está corroborada. Ahora bien, para pasar de «corroborada» a «fiable», usted examinó el poder explicativo de la teoría. Así que lo que ha demostrado es que lo que podríamos denominar el «principio de buscar mejores explicaciones», junto con algunas observaciones, sí, sí, y argumentaciones, *implica* que el futuro se parecerá, en muchos aspectos, al pasado. ¡Y esto es un principio de inducción! Si su «principio de explicación» implica un principio de inducción, lógicamente, es un principio de inducción. De modo que el inductivismo es cierto, después de todo, y el principio de inducción debe ser postulado, explícita o implícitamente, para que podamos predecir el futuro.

DAVID: ¡Oh, cielos! ¡El inductivismo es una enfermedad realmente virulenta! Parecía haber remitido durante unos instantes, pero veo que vuelve con más fuerza que antes.

CRIPTOINDUCTIVISTA: ¿Justifica también el racionalismo popperiano los argumentos *ad hominem?* Lo pregunto por pura curiosidad, ¿sabe?

DAVID: Le pido disculpas. Permítame ir directamente a la sustancia de lo que ha dicho. Sí, he justificado una afirmación sobre el futuro. Dice usted que ello implica que «el futuro se parezca al pasado». Bueno, superficialmente, sí, en la medida en que *cualquier* teoría sobre el futuro afirmará que se parece en algún sentido al pasado. Pero esa inferencia de que el futuro se parece al pasado no es el principio de inducción perseguido, ya que nunca podríamos derivar ni justificar teoría o predicción alguna a partir de ella. Por ejemplo, no la podríamos utilizar para diferenciar su teoría de la gravedad de la dominante, puesto que ambas dicen, cada una a su manera, que el futuro se parece al pasado.

CRIPTOINDUCTIVISTA: ¿No podríamos deducir del «principio de explicación» una forma de principio de inducción que *pudiera* ser utilizada para la selección de teorías? ¿Qué le parecería: «Si una anomalía inexplicada no sucede en el pasado, no es probable que suceda en el futuro.»?

DAVID: No. Nuestra justificación no depende de que una determinada anomalía suceda en el pasado. Está relacionada con el hecho de si hay o no una explicación para su existencia.

CRIPTOINDUCTIVISTA: Muy bien, permítame formularlo más cuidadosamente: «Si en el presente no existe ninguna teoría explicativa que prediga que una determinada anomalía sucederá en el futuro, no es probable que dicha anomalía suceda en el futuro.»

DAVID: Esto podría muy bien ser cierto. Yo, sin ir más lejos, creo que lo es. No obstante, no es un enunciado semejante a «el futuro se parecerá probablemente al pasado». Es más, al tratar de hacer que se le asemeje tanto como sea posible, lo ha limitado a los casos «en el presente» y «en el futuro», y al caso de ser una «anomalía». Pero sería igual de cierto sin estas limitaciones. Es, simplemente, una afirmación general sobre la eficacia de la argumentación. Resumiendo, si no existe argumentación en favor de un postulado, no es fiable. Pasado, presente o futuro. Anomalía o no. Y punto.

CRIPTOINDUCTIVISTA: Ya.

DAVID: En los conceptos «argumentación racional» o «explicación» nada relaciona el futuro con el pasado de un modo especial. No se postula, ni mucho menos, el «parecido» de una cosa con la otra, y no serviría de nada hacerlo. El sentido superficial en que el concepto de «explicación» implica que el futuro «se parece al pasado» no incluye nada específico sobre el futuro, de modo que no puede tratarse de un principio de inducción. No hay ningún principio de inducción. No hay ningún proceso de inducción. Nadie los utiliza nunca, ni nada que se les parezca. No existe ya el problema de la inducción. ¿Está claro ahora?

CRIPTOINDUCTIVISTA: Sí. Tenga un poco de paciencia, por favor, mientras reajusto mi concepción del mundo.

DAVID: Para ayudarle a hacerlo, le aconsejo que considere su «teoría de la gravedad» alternativa más atentamente.

CRIPTOINDUCTIVISTA: ¿Qué?

DAVID: Estamos de acuerdo en que su teoría consiste, básicamente, en una teoría de la gravedad, la dominante, restringida por una predicción inexplicada sobre mí. Afirma que planearía sin ser sostenido. «Sin ser sostenido» significa «sin que ninguna fuerza ascendente actúe» sobre mí, lo que sugiere que sería inmune a la «fuerza» de la gravedad, que, de otro modo,

tiraría de mí hacia abajo. Pero, según la teoría general de la relatividad, la gravedad no es una fuerza, sino una manifestación de la curvatura del espacio-tiempo. Esta curvatura explica por qué objetos no sostenidos, como yo, si caigo, y la Tierra, se acercan con el tiempo. Por lo tanto, y a la luz de la física moderna, su teoría, presumiblemente, afirma que *hay* una fuerza ascendente que actúa sobre mí, indispensable para mantenerme a una distancia constante de la Tierra. Pero ¿de dónde procede esa fuerza, y cómo se comporta? Por ejemplo, ¿qué es una «distancia constante»? Si la Tierra se moviese hacia abajo, ¿respondería yo instantáneamente para mantener mi altura con respecto a ella, lo que requeriría una velocidad de comunicación superior a la de la luz y contradice otro principio de la relatividad, o me llegaría primero, a la velocidad de la luz, la información acerca de dónde se encuentra la Tierra? En este caso, ¿qué transportaría esta información? ¿Es una nueva clase de onda emitida por la Tierra? Si es así, ¿por qué ecuaciones se rige? ¿Transporta energía? ¿Cuál es su comportamiento, de acuerdo con la mecánica cuántica? ¿Se trata, tal vez, de que respondo de un modo especial a las ondas existentes, tales como la luz? En este caso, ¿desaparecería la anomalía si fuese colocada una barrera opaca entre la Tierra y yo? Pero ¿no es la Tierra, fundamentalmente, opaca? ¿Y dónde empieza la «Tierra»: qué define la superficie sobre la cual se supone que «planeo»?

CRIPTOINDUCTIVÍSTA: ¡Difícil pregunta!

DAVID: Sí. Y no lo es menos la de dónde empiezo yo y qué me define. Si me agarro a algo pesado, ¿planeará también? De ser así, a los aviones en que he volado habrían podido fallarles los motores sin que hubiese ocurrido ninguna desgracia. ¿Y qué significa «agarrarse»? ¿Caería el avión en picado si levantase el codo del brazo de mi asiento? Y, si mi capacidad para descender planeando no alcanza a las cosas a las que me agarre, ¿qué hay de mis ropas? ¿Tirará de mí su peso y hará que al fin me estrelle y me mate, después de todo, si salto por encima de la barandilla? ¿Y qué hay de mi última comida?

CRIPTOINDUCTIVISTA: ¡Ciertamente!

DAVID: Podría seguir así *ad infinitum*. Lo cierto es que, cuanto más consideremos las implicaciones de la anomalía que propone, más preguntas sin contestar encontraremos. No se trata sólo de que su teoría esté incompleta. Estas preguntas son *dilemas*. Cualquiera que sea su respuesta, crean nuevos

problemas al invalidar explicaciones satisfactorias existentes para otros fenómenos.

CRIPTOINDUCTIVISTA: ¿Y bien?

DAVID: Por lo tanto, su postulado adicional no es sólo superfluo, sino positivamente erróneo. En general, las teorías inanes pero no refutadas, que se puede sacar uno de la manga encajan, *grosso modo*, en dos categorías. Por un lado, la de las teorías que postulan entidades inobservables, tales como la existencia de partículas que sólo interactúan entre ellas. Pueden ser rechazadas, porque no solucionan nada. Se les aplica la navaja de Occam, de hecho. Y, por otro lado, la de las teorías que, como la suya, predicen anomalías observables, pero inexplicadas. Pueden ser rechazadas, porque, *además* de no solucionar nada, invalidan soluciones existentes. No se trata, me apresuro a decir, de que entren en conflicto con las observaciones existentes, sino de que privan de capacidad explicativa a teorías generalmente aceptadas al afirmar que sus predicciones tienen excepciones, pero sin explicar el porqué. Usted no puede decir, simplemente, «la geometría espaciotemporal hace que se atraigan mutuamente los objetos que no son sostenidos, excepto cuando uno de ellos es David; en este caso no los afecta». O bien la curvatura del espacio—tiempo es la explicación de la gravedad, o no. Compare, si no, su teoría con la perfectamente legítima afirmación de que una pluma descenderá planeando con suavidad porque el aire ejercerá sobre ella una fuerza ascensional suficiente. Esta afirmación es consecuencia de la teoría existente que explica qué es el aire, de modo que no crea ningún problema nuevo, mientras que la suya sí.

CRIPTOINDUCTIVISTA: Ya veo. ¿Me ayudará a reajustar mi concepción del mundo?

DAVID: Bueno, ¿ha leído mi libro *La estructura de la realidad*?

CRIPTOINDUCTIVISTA: Pienso hacerlo, pero, de momento, la ayuda que le pido concierne a una dificultad muy específica.

DAVID: Adelante.

CRIPTOINDUCTIVISTA: La dificultad es la siguiente: si repaso mentalmente la discusión que estamos manteniendo, no me cabe la menor duda de que su predicción de lo que sucedería si usted o yo saltásemos desde esta torre no se deriva de ninguna hipótesis inductiva como «el futuro se parece al pasado». Pero si doy un paso atrás y considero en su totalidad la lógica de la situación, me

temo que sigo sin entender cómo puede ser así. Consideremos la materia prima de la argumentación. Inicialmente, asumí que las observaciones pasadas y la lógica deductiva eran toda nuestra materia prima. Luego admití que el problema-situación en curso también es relevante, puesto que sólo necesitamos justificar que nuestra teoría es más fiable que sus rivales. Luego tuve que tomar en consideración que vastas clases de teorías pueden ser descartadas, simplemente, mediante la argumentación, puesto que son malas explicaciones, y también que, en consecuencia, los principios de racionalidad pueden ser incluidos entre nuestras materias primas. Lo que no llego a entender es cómo puede nacer la justificación de futuras predicciones de esta masa de materia prima, formada por observaciones *pasadas*, el problema-situación *presente* y principios *atemporales* de lógica y racionalidad, ninguno de los cuales justifica inferencias del pasado al futuro. Aquí parece haber un vacío lógico. ¿Estamos haciendo una asunción oculta en alguna parte?

DAVID: No, no hay ningún vacío lógico. Lo que usted denomina «nuestra materia prima» incluye, sin duda, aserciones sobre el futuro. Las mejores teorías existentes, que no pueden ser desdeñadas a la ligera, puesto que son soluciones de problemas, contienen predicciones sobre el futuro. Estas predicciones no pueden ser separadas del resto del contenido de las teorías, como intentaba hacer usted, ya que ello menoscabaría su capacidad explicativa. Cualquier nueva teoría propuesta debe, por lo tanto, *o bien* ser coherente con las teorías existentes, lo que tiene implicaciones respecto de lo que la nueva teoría pueda decir sobre el futuro, *o bien* contradecir algunas de las teorías existentes, pero afrontando los nuevos problemas planteados y ofreciendo explicaciones alternativas, lo cual reduce, a su vez, lo que puedan decir sobre el futuro.

CRIPTOINDUCTIVISTA: De modo que no tenemos ningún principio de razonamiento que diga que el futuro se parecerá al pasado, pero sí teorías que lo dicen. ¿Quiere decir eso que disponemos de teorías que implican una forma limitada del principio de inducción?

DAVID: No. Nuestras teorías, simplemente, afirman algo sobre el futuro. De un modo superficial, toda teoría sobre el futuro implica que éste «se parecerá al pasado» en algunos aspectos. Pero sólo podemos saber en qué aspectos dice que lo hará cuando disponemos de ella. Por esa misma regla de tres, se podría decir

que, puesto que nuestras teorías afirman que ciertas características de la realidad se mantienen a través del *espacio*, implican un «principio espacial de inductivismo» en el sentido de que «lo cercano se parece a lo lejano». Debo hacer hincapié en lo siguiente: en todos los sentidos prácticos de la palabra «parecerse», nuestras teorías actuales dicen que el futuro *no* se parecerá al pasado. Por ejemplo, el Big Crunch cosmológico, la contracción del universo hasta volver a su punto de partida, es un acontecimiento que algunos cosmólogos predicen, pero, desde un punto de vista físico, la posibilidad de que ocurra es tan remota en un hipotético futuro como en la época actual. Las propias leyes que sirven de base para esa predicción ya no le serían aplicables.

CRIPTOINDUCTIVISTA: Estoy convencido de ello. Déjeme ensayar una última argumentación. Hemos visto que las predicciones futuras pueden ser justificadas recurriendo a los principios de racionalidad, pero ¿qué justifica a éstos? No son, después de todo, verdades de lógica pura, de modo que existen dos posibilidades: o bien no están justificados, en cuyo caso las conclusiones de ellos derivadas tampoco lo están, o bien los justifica alguna razón todavía desconocida. En ambos casos, falta una justificación. Ya no sospecho que éste sea el problema de la inducción disfrazado. Sin embargo, al refutar de modo tan rotundo el problema de la inducción, ¿no habremos destapado otro problema fundamental, también de falta de justificación, que aquél ocultaba?

DAVID: ¿Qué justifica los principios de racionalidad? La argumentación, como de costumbre. ¿Qué justifica, por ejemplo, nuestra confianza en las leyes de *deducción*, a pesar de que todo intento de justificarlas lógicamente nos conduzca de manera irremediable a la tautología o a una regresión infinita? La justifica que prescindir de ellas no mejora ninguna explicación.

CRIPTOINDUCTIVISTA: No parece una base muy firme para la lógica pura.

DAVID: No es una base firme, ni deberíamos esperar que lo fuese, puesto que el razonamiento lógico es un proceso físico, como el razonamiento científico, y, por consiguiente, intrínsecamente falible. Las leyes de la lógica no son evidentes por sí mismas. Hay quienes, como los «intuicionistas» matemáticos, no están de acuerdo con las leyes convencionales de deducción, es decir, las «reglas lógicas de inferencia». Comento su insólita concepción del mundo en el capítulo 10 de *La estructura de la*

realidad. No se puede *probar* que estén equivocados, pero *argumentaré* que lo están, y estoy seguro de que usted estará de acuerdo en que mi argumentación justifica esa conclusión.

CRIPTOINDUCTIVISTA: Así pues, ¿no cree que exista un «problema de la inducción»?

DAVID: No. No creo que sea problemática ninguna de las formas habituales de justificación de conclusiones en ciencia, filosofía o matemáticas. Sin embargo, es un *hecho* interesante que el universo físico admita procesos que crean conocimiento sobre él, al igual que sobre otras cosas. Podemos tratar de explicar razonablemente este hecho del mismo modo en que explicamos otros hechos físicos, es decir, mediante teorías explicativas. Como podrá ver en el capítulo 6 de *La estructura de la realidad*, en mi opinión, el principio de Turing es la teoría indicada en este caso. Dice que es posible construir un generador de realidad virtual cuyo repertorio incluya todos los entornos físicamente posibles. Si el principio de Turing es una ley de física, como he argumentado, no deberíamos sorprendernos de descubrir que podemos elaborar teorías acertadas sobre la realidad, ya que esto no es más que realidad virtual en acción. Al igual que el hecho de que las máquinas de vapor sean posibles es consecuencia directa de los principios de la termodinámica, el hecho de que el cerebro humano sea capaz de crear conocimiento es consecuencia directa del principio de Turing.

CRIPTOINDUCTIVISTA: Pero ¿cómo podemos estar seguros de que el principio de Turing es *cierto*?

DAVID: No podemos, por descontado... ¡Ah! Ahora tiene miedo de que, si no podemos justificar el principio de Turing, nos encontremos de nuevo con que no podemos justificar nuestra confianza en las predicciones científicas, ¿verdad?

CRIPTOINDUCTIVISTA: Bueno.. Sí.

DAVID: ¡Pero si se trata de una cuestión completamente distinta! Ahora estamos considerando un *hecho* evidente relacionado con la realidad física, el de que puede hacer predicciones fiables acerca de sí misma. Tratamos de explicar ese hecho, de situarlo dentro del mismo marco que los demás hechos que conocemos. He sugerido que puede estar implicada alguna ley de la física. Pero si me equivocase, incluso si a todo el mundo le resultase imposible llegar a explicar esta notable propiedad de la realidad, ello no menoscabaría ni un ápice la justificación de ninguna teoría científica, porque no empeoraría ni un ápice las

explicaciones que proporcionan dichas teorías.

CRJPTOINDUCTIVISTA: Ahora sí que mis argumentaciones se han agotado. Intelectualmente, me declaro convencido. Sin embargo, debo confesar que siento aún lo que sólo se me ocurre describir como una «duda emocional».

DAVID: Quizás le sea de ayuda un último comentario, no sobre ninguna de las argumentaciones específicas que ha planteado, sino sobre un error que parece subyacer en muchas de ellas. Usted sabe que se trata de un error, pero quizás no se ha percatado de sus consecuencias para su concepción del mundo. Es probable que sea el origen de su «duda emocional».

CRJPTOINDUCTIVISTA: Adelante.

DAVID: Es un error sobre la naturaleza intrínseca de la argumentación y la explicación. Parece usted asumir que las argumentaciones y las explicaciones, tales como las que justifican obrar de acuerdo con una determinada teoría, revisten la forma de pruebas matemáticas y proceden de asunciones a conclusiones. Busca usted la «materia prima», los axiomas, de la que se derivan nuestras conclusiones, los teoremas. Hay, ciertamente, una estructura lógica de esa clase asociada a cada argumentación o explicación que resulta ser válida, pero el proceso de argumentación no empieza por los «axiomas» y termina en la «conclusión». Empieza más bien en el centro, con una versión cargada de inconsistencias, vacíos, ambigüedades e irrelevancias. Todos estos defectos son criticados. Se hacen intentos para reemplazar las teorías defectuosas. Las teorías que son criticadas y reemplazadas contienen, habitualmente, algunos de los «axiomas». Por ello es un error asumir que una argumentación empieza con las teorías que eventualmente le sirven de «axiomas» o es justificada por ellas. La argumentación termina, siempre de modo provisional, cuando parece haber demostrado que la explicación asociada es satisfactoria. Los «axiomas» adoptados no constituyen creencias finales e incuestionables. Son teorías provisionales, que sólo intentan explicar.

CRIPTOINDUCTIVISTA: Ya. La argumentación no es de la misma especie que la deducción o que la inexistente inducción. No se basa en nada ni se justifica por nada. *Y tampoco lo necesita*, puesto que su objetivo es resolver problemas, demostrar que un determinado problema queda resuelto por una determinada explicación.

DAVID: ¡Bienvenido al club!

EX INDUCTIVISTA: ¡Me he sentido tan seguro todos estos años con mi Problema con mayúscula! ¡Me sentía tan superior, tanto a los antiguos inductivistas como al arrogante Popper!

Y durante todo ese tiempo, sin saberlo, ¡era un criptoinductivista! El inductivismo es, en efecto, una enfermedad. Te ciega. DAVID: No sea tan duro consigo. Ya está curado. ¡Ojalá sus compañeros de sufrimientos estuviesen tan dispuestos como usted a ser curados por la simple argumentación!

EX INDUCTIVISTA: Pero ¿cómo he podido estar tan ciego? ¡Y pensar que llegué a postular la candidatura de Popper para el Premio de las Ideas Ridiculas cuando, en realidad, había resuelto el problema de la inducción! ¡Oh, *mea culpa*! ¡Que Dios nos perdone, porque hemos quemado a un santo! ¡Me siento tan avergonzado! No veo otra salida que saltar por encima de la barandilla.

DAVID: No creo que sea necesario. Nosotros, los popperianos, preferimos que mueran las teorías. Arroje el *inductivismo* al vacío, en vez de saltar usted.

TERMINOLOGÍA

Criptoinductivista. Alguien que cree que la invalidez del razonamiento inductivista plantea el serio problema filosófico de cómo justificar la confianza en las teorías científicas.

En el próximo capítulo consideraremos la cuarta vía, la teoría de la evolución, que responde a la pregunta: «¿Qué es la vida?»

8. LA TRASCENDENCIA DE LA VIDA

Desde la antigüedad hasta, aproximadamente, el siglo XIX se daba por sentado que era necesaria la existencia de una fuerza o factor animante especial para que la materia de los organismos vivos se comportara de modo tan notoriamente diferente que el resto de la materia. Se postulaba, pues, la existencia de dos clases de materia, *animada e inanimada*, dotadas de propiedades físicas distintas por completo. Consideremos un organismo vivo, por ejemplo un oso. Una fotografía de un oso se parece a un oso en determinados aspectos. También se parecen a un oso otros objetos inanimados, tales como el cadáver de un oso e incluso, de un modo más limitado, la constelación de la Osa Mayor. Pero sólo la materia animada puede perseguirnos a través del bosque mientras intentamos escapar zigzagueando entre los árboles y destrozarnos si nos atrapa. Las cosas inanimadas nunca harían algo tan premeditado... o, al menos, así lo creían los antiguos. Por supuesto, no habían visto nunca un cohete dirigido.

Para Aristóteles y otros filósofos de la antigüedad, la facultad más espectacular de la materia animada era su capacidad de moverse voluntariamente. Consideraban que cuando la materia inanimada, como una piedra, queda en reposo, no vuelve a moverse a menos de ser golpeada por algo. La materia animada, en cambio, puede estar en reposo, como un oso en hibernación, y reanudar luego el movimiento sin necesidad de ser golpeada. Gracias a la ciencia moderna, podemos fácilmente detectar fallos en estas generalizaciones. Por ejemplo, la idea de «movimiento voluntario» parece ahora errónea: sabemos que el oso se despierta a causa de los procesos electroquímicos de su cuerpo. Éstos pueden ser iniciados por «golpes» externos, tales como cambios de temperatura, o por su reloj biológico interno, que utiliza lentas reacciones químicas para seguir el paso del tiempo.

Las reacciones químicas son movimientos de átomos, de modo que el oso nunca se encuentra en completo reposo. Por otro lado, un núcleo de uranio, que, sin duda, no está vivo, puede permanecer estable durante miles de millones de años hasta que, sin ningún estímulo exterior, un buen día se desintegra de manera súbita y violenta. Así pues, el contenido nominal de la idea de Aristóteles no es válido en la actualidad. No obstante, el Estagirita tenía muy claro algo importante, en lo que, en cambio, se

equivocan muchos de los modernos pensadores. Al tratar de asociar la vida con un concepto físico básico (aunque fuera el de movimiento, en lo que estaba equivocado), reconocía que ésta es un fenómeno fundamental de la naturaleza.

Un fenómeno es «fundamental» cuando de su comprensión depende que se comprenda de manera más profunda el mundo. Por supuesto, no existe unanimidad acerca de qué aspectos del mundo vale la pena comprender y, por consiguiente, acerca de lo que es profundo o fundamental. Algunos dirán que el amor es el fenómeno más fundamental del mundo. Otros piensan que, si se aprenden ciertos textos sagrados de memoria, se entiende todo lo que vale la pena entender. La comprensión de la que hablo se expresa en leyes de física y en principios de lógica y filosofía. Una comprensión «más profunda» es aquella que resulta más general, implica menos conexiones entre verdades superficialmente diversas, explica más y deja menos asunciones sin explicar. Los fenómenos más fundamentales intervienen en la explicación de muchísimos fenómenos, pero son explicados únicamente por leyes y principios básicos.

Al contrario de lo que ocurre con la gravitación, que es de los más fundamentales, no todos los fenómenos de esta clase tienen grandes efectos físicos. Los efectos directos de la interferencia cuántica, como las conformaciones fantasmas descritas en el capítulo 2, son tan reducidos que resulta incluso difícil su detección inequívoca. Sin embargo, hemos visto que la interferencia cuántica es un fenómeno fundamental. Sólo si lo comprendemos podremos entender el hecho básico acerca de la realidad física, es decir, la existencia de universos paralelos.

Resultaba evidente para Aristóteles que la vida es fundamental teoréticamente y tiene importantes efectos físicos. Como veremos, tenía razón, pero ello resultaba obvio para él por razones erróneas; por ejemplo, las supuestamente distintas propiedades mecánicas de la materia animada y la dominación de la superficie de la Tierra por los procesos vitales. Aristóteles pensaba que el universo consistía, básicamente, en lo que ahora denominamos la biosfera (la región que contiene la vida) de nuestro planeta, con algunas regiones adicionales (las esferas celestiales y el interior de la Tierra) clavadas, por así decirlo, encima y debajo de ese núcleo central. Cuando la biosfera es considerada el componente principal del cosmos, es natural pensar que los animales y las plantas son, al menos, tan importantes

como las rocas y las estrellas en el gran esquema de las cosas, en especial si se sabe muy poco de física o biología. La ciencia moderna ha llegado a una conclusión casi totalmente opuesta. La revolución copernicana convirtió a la Tierra en mero satélite de un Sol central e inanimado. Los posteriores descubrimientos en física y astronomía no sólo demostraron que el universo es vasto en comparación con la Tierra, sino que está perfectamente descrito mediante leyes globales que no hacen alusión a la vida para nada. La teoría de la evolución de Charles Darwin explicó el origen de la vida en términos que no requerían una física especial, y desde entonces son muchos los mecanismos detallados de la vida que se han ido descubriendo, sin que en ninguno de ellos se haya hallado tampoco una física especial.

Esos éxitos espectaculares de la ciencia, así como la gran generalidad de la física newtoniana y, en particular, de la física subsiguiente, contribuyeron en buena medida a hacer atractivo el reduccionismo. Puesto que la fe en la verdad revelada había demostrado ser incompatible con la racionalidad (que requiere apertura a la crítica), muchas personas anhelaban hallar una base definitiva para las cosas en la que pudiesen creer. Si bien aún no disponían de una «teoría total» reduc-tiva en la que creer, no por ello dejaban de aspirar a tenerla. Se dio por sentado que una jerarquía reduccionista de las ciencias, basada en la física subatómica, era inherente a la visión científica del mundo, y, en consecuencia, quienes la criticaban no eran más que pseudocientíficos y seres rebeldes que no querían aceptar la autoridad de la ciencia. Por ello, para cuando yo estudié biología en la escuela, el status de esta materia se había convertido en el contrario del que Aristóteles consideraba evidente. La vida ya no era considerada fundamental. Incluso el término «estudio de la naturaleza» —en el sentido de biología— se había convertido en un anacronismo. Fundamentalmente, la naturaleza era física. Aunque simplifico y exagero un poco, la visión dominante era la que expongo a continuación. La física tenía una ramificación, la química, que estudiaba la interacción de los átomos. La química tenía a su vez una ramificación, la química orgánica, que estudiaba las propiedades de los compuestos del elemento carbono. La química orgánica tenía, asimismo, una ramificación, la biología, que estudiaba los procesos químicos que denominamos vida. Sólo porque resulta que *somos* uno de esos procesos era interesante para nosotros esa remota ramificación de una materia

fundamental. La física, en cambio, se consideraba, evidentemente, importante por derecho propio, ya que el universo todo —incluida la vida— se conforma a sus principios.

Mis condiscípulos y yo debimos aprendernos de memoria una serie de «características de las cosas vivas». Éstas eran meramente descriptivas y hacían escasa referencia a conceptos fundamentales. Por supuesto, la locomoción o movimiento era una de ellas —un mal definido eco de la idea aristotélica—, y tampoco faltaban la *respiración y* la *excreción.* Estaban también la *reproducción*, el *crecimiento y* una de nombre inolvidable, la *irritabilidad*, en el sentido de que, «si lo golpeas, devuelve el golpe». Lo que les faltaba a esas supuestas características de la vida en exactitud científica *y* profundidad no lo compensaba la precisión terminológica. Como había dicho el doctor Johnson, todo objeto real es «irritable». Por otro lado, los virus no respiran, crecen, excretan o se mueven (a menos de ser «golpeados»), y, sin embargo, están bien vivos. Asimismo, los seres humanos estériles no se reproducen, y no por ello están menos vivos.

La razón por la que tanto la concepción de Aristóteles como la de los libros de texto de mi época ni siquiera eran capaces de hacer una buena distinción taxonómica entre cosas vivas y no vivas —y no hablemos de cualquier distinción más profunda—, es que ninguna de las dos se percataba de qué es, en realidad, una entidad viva (un error más perdonable en tiempo de Aristóteles, ya que entonces todo el mundo opinaba como él). La biología moderna no trata de definir la vida mediante algún atributo físico o alguna sustancia —como una «esencia» vital—, características de los que tan sólo la materia viva está dotada. No se espera ya encontrar esencia alguna, puesto que sabemos que la «materia animada» —la materia en forma de organismos vivos— no constituye la base de la vida. Es meramente uno de los efectos de la vida, cuya base es molecular. Es un hecho que existen determinadas moléculas que hacen que determinados entornos las copien.

Estas moléculas reciben el nombre de *replicantes.* Más generalmente, un replicante es una entidad que hace que ciertos entornos la copien. No todos los replicantes son biológicos, ni todos son moléculas. Un programa informático autocopiante (como un virus informático), por ejemplo, es también un replicante. Un buen chiste es igualmente un replicante, puesto que provoca que los que lo han escuchado lo mencionen para repetirlo ante otras

audiencias. Richard Dawkins ha acuñado el término *meme* para designar los replicantes que son ideas humanas, como los chistes. Pero toda la vida sobre la Tierra se basa en replicantes que son moléculas. Se llaman *genes*, y la biología es el estudio de sus orígenes, estructura y funcionamiento, así como de sus efectos sobre otra materia. En la mayoría de los organismos, un gen consiste en una secuencia de moléculas más pequeñas, de las que hay cuatro clases distintas, unidas formando una cadena. Los nombres de las moléculas componentes (adenina, citosina, guanina y timina) se utilizan habitualmente abreviados: A, C, G y T. El nombre químico abreviado de una cadena con cualquier número de moléculas A, C, G y T, dispuestas en cualquier orden, es ADN o ácido desoxirribonucleico.

Los genes son verdaderos programas informáticos, escritos como secuencias de los símbolos A, C, G y T, en un lenguaje estándar que se denomina *código genético*, el cual, con muy ligeras variaciones, es común a toda la vida sobre la Tierra. (Algunos virus se basan en un tipo de molécula relacionada, el ARN o ácido ribonucleico, mientras que los priones son, en cierto sentido, proteínas moleculares autorreplicantes.) Determinadas estructuras operan en el interior de las células como ordenadores que ejecutan esos programas genéticos. Dicha ejecución consiste en la obtención de ciertas moléculas (proteínas) a partir de moléculas más simples (aminoácidos), en determinadas condiciones externas. La secuencia «ATG», por ejemplo, es una instrucción para incorporar el aminoácido metionina a la molécula de proteína que se está obteniendo.

De modo característico, un gen es «activado» químicamente en determinadas células del cuerpo, a las que instruye para fabricar la correspondiente proteína. Por ejemplo, la hormona insulina, responsable del control del nivel de azúcar en la sangre de los vertebrados, es una de esas proteínas. Si bien el gen para su fabricación está presente en casi todas las células del cuerpo, sólo es activado en determinadas células especializadas del páncreas, y ello únicamente cuando es necesario. A nivel molecular, eso es todo lo que puede hacer un gen: programar a su ordenador celular para fabricar un determinado compuesto químico. Sin embargo, los genes consiguen actuar como replicantes porque esos programas químicos de bajo nivel se suman, mediante capas y más capas de complejos controles y retroalimenta-dones, hasta convertirse en complicadísimas

instrucciones de alto nivel. Conjuntamente, el gen de la insulina y los genes involucrados en su activación y desactivación constituyen un completo programa para la regulación del nivel de azúcar en el torrente sanguíneo.

Del mismo modo, existen genes que contienen instrucciones específicas que determinan cómo y cuándo deben ser copiados, ellos y otros genes, así como instrucciones para la fabricación de nuevos organismos de la misma especie, incluyendo los ordenadores moleculares que ejecutarán todas esas instrucciones generación tras generación. Hay también instrucciones acerca de cómo debe responder el organismo en su totalidad ante los estímulos, como por ejemplo, cuándo ha de cazar, comer, copular, luchar o huir. Y así sucesivamente.

Un gen puede actuar como replicante tan sólo en determinados entornos. Por analogía con un «nicho» ecológico (el conjunto de entornos en los que un organismo puede sobrevivir y reproducirse), utilizaré también el término *nicho* para referirme al conjunto de todos los posibles entornos en que un determinado replicante podría hacer que lo copiasen. El nicho de un gen de la insulina incluye entornos en los que está ubicado en el núcleo de una célula en compañía de otros genes, y la célula, por su parte, se encuentra adecuadamente situada en el interior de un organismo funcional, en un hábitat adecuado para su vida y su reproducción. Pero hay también otros entornos capaces de copiar el gen de la insulina, como los laboratorios de biotecnología, en los que bacterias son alteradas genéticamente para incorporarles el gen. Dichos entornos también forman parte del nicho del gen, al igual que otros posibles entornos, infinitamente numerosos, todos ellos muy distintos de aquel en que evolucionó.

No todo lo que puede ser copiado es un replicante. Éste *causa* que su entorno lo copie, es decir, contribuye causalmente a su propia copia. (Mi terminología difiere ligeramente de la utilizada por Daw-kins, quien denomina replicante a todo aquello que es copiado, cualesquiera que sea la razón. Lo que yo denomino replicante, Dawkins lo denominaría replicante *activo.)* Volveré más adelante sobre el significado general de contribuir causalmente a algo, pero lo que quiero decir aquí es que la presencia del replicante y su forma física específica *marcan la diferencia* entre si va a ser copiado o no. En otras palabras, el replicante es copiado si está presente, pero si fuese reemplazado por otro objeto, incluso muy similar, no lo sería. El gen de la insulina, por ejemplo,

provoca solamente un pequeño paso en el enormemente complicado proceso de su propia replicación (ese proceso es el ciclo vital completo del organismo). Pero la inmensa mayoría de las variantes de dicho gen no podrían instruir a las células para elaborar un compuesto químico capaz de realizar las funciones de la insulina. Si los genes de la insulina de un organismo fuesen reemplazados por moléculas ligeramente distintas, dicho organismo moriría (a menos que fuese mantenido en vida por otros medios), y, por consiguiente, no podría reproducirse, de modo que las moléculas en cuestión no podrían ser copiadas. Así pues, que la copia tenga lugar o no es algo extraordinariamente sensible a la forma física del gen de la insulina. Su presencia en la forma y el lugar adecuados *marca la diferencia* entre si se realizará la copia o no, lo que lo convierte en un replicante, si bien existen innumerables otras causas que contribuirán también a su replicación.

Junto con los genes, en el ADN de la mayoría de los organismos están presentes secuencias *aleatorias* de A, C, G y T, denominadas a veces secuencias *basura del ADN.* Son también copiadas y transmitidas a la descendencia. Sin embargo, si semejante secuencia es sustituida por casi cualquier otra de parecida longitud sigue siendo copiada, por lo que podemos deducir que la copia de estas secuencias no depende de su forma física específica. A diferencia de los genes, las secuencias basura no son programas. Si tienen alguna función (y no se sabe aún que la tengan), no puede ser la de transmitir información de ninguna clase. Si bien son copiadas, no contribuyen causalmente a su copia, por lo que no pueden ser consideradas replicantes.

En realidad, esto es una exageración. Todo lo que ha sido copiado debe haber hecho alguna contribución causal a su copia. Las secuencias basura, por ejemplo, están hechas de ADN, lo que permite al ordenador celular copiarlas. No puede copiar moléculas que no sean de ADN. No obstante, no suele resultar ilustrativo considerar replicante a algo que tenga tan sólo una pequeña contribución causal a su propia réplica. Estrictamente hablando, ser replicante es cuestión de grado. Definiré el *grado de adaptación* de un replicante a un determinado entorno como el grado en el que el replicante contribuye causalmente a su propia réplica en dicho entorno. Cuando un replicante está bien adaptado a la mayoría de los entornos de un nicho, podemos definirlo como bien adaptado al nicho. Hemos visto que el gen de la insulina está

altamente adaptado a su nicho. Las secuencias basura tienen un grado despreciable de adaptación en comparación con el del gen de la insulina o cualquier gen fidedigno, pero están mucho más adaptadas a dicho nicho que la mayoría de las demás moléculas.

Téngase en cuenta que, para cuantificar grados de adaptación, debemos tomar en consideración no sólo el replicante en cuestión, sino también el abanico de sus posibles variantes. Cuanto más sensible a la estructura física precisa del replicante sea el proceso de copia en un determinado entorno, más adaptado estará dicho replicante a ese entorno. Para los replicantes altamente adaptados (los únicos que merecen esa denominación) debemos considerar sólo ligeras variaciones, puesto que si se dieran variaciones más amplias ya no serían replicantes. Consideraremos, pues, la sustitución del replicante por objetos más o menos semejantes. Para cuantificar el grado de adaptación de un replicante a un nicho, debemos tener en cuenta su grado de adaptación a todos los entornos del nicho. Deberemos considerar, pues, tanto las variantes del entorno como las del replicante. Si la mayoría de las variantes del replicante no consiguen causar que la mayoría de los entornos las copien, se desprende que la forma del replicante en cuestión es la causa determinante de su propia copia en aquel nicho, que es lo que queremos decir con que está altamente adaptado a él. Por otro lado, si la mayoría de las variantes del replicante son copiadas en la mayoría de los entornos del nicho, ello significa que la forma de ese replicante no incide en la copia, que se producirá en cualquier caso. En consecuencia, el replicante tiene escasa contribución causal a su propia copia y, por lo tanto, no está altamente adaptado a este nicho.

Así pues, el grado de adaptación de un replicante depende no sólo de lo que haga en su entorno concreto, sino también de lo que *harían* múltiples objetos —la mayoría de los cuales no existen— en un amplio número de entornos distintos del concreto. Ya nos hemos encontrado antes con esta curiosa propiedad. Por ejemplo, la fidelidad de una representación en realidad virtual depende no sólo de las respuestas efectivas de la máquina a las acciones reales del usuario, sino también de las respuestas que no tiene, llegado el caso, a acciones que el usuario no hace de hecho. Esta similitud entre los procesos vitales y la realidad virtual no es pura coincidencia, como explicaré en breve.

El factor más importante que determina el nicho de un gen es, normalmente, que su replicación depende de la presencia de

otros genes. La réplica de un gen de insulina de un oso, por ejemplo, depende no sólo de la presencia en el cuerpo del oso de todos sus restantes genes, sino también de la presencia en el medio exterior de otros genes de otros organismos. Los osos no pueden sobrevivir sin comida, y los genes que fabrican esa comida existen solamente en otros organismos.

A menudo, distintas clases de genes que se necesitan mutuamente para su replicación conviven unidos en largas cadenas del ADN: el ADN de un *organismo.* Un organismo es la clase de cosa —un animal, una planta o un microbio, por ejemplo— que, en términos cotidianos, consideramos viva. Sin embargo, de lo que he dicho anteriormente se desprende que «viva» es, en el mejor de los casos, una manera de hablar cuando la aplicamos a cualquier parte de un organismo que no sea su ADN. *Un organismo no es un replicante*, sino una parte del entorno de replicantes; habitualmente la parte más importante, después de los demás genes. El resto del entorno lo constituyen la clase de hábitat ocupado por el organismo (por ejemplo, cimas montañosas o profundidades oceánicas) y el estilo de vida particular (caza, microfagia,[6]etcétera) que, dentro de ese hábitat, capacita a dicho organismo para vivir el tiempo suficiente a fin de que sus genes puedan ser replicados.

En el lenguaje cotidiano, decimos que los organismos «se reproducen». No olvidemos que ésta era una de las supuestas «características de las cosas vivas». En otras palabras, pensamos que los organismos son replicantes, pero esto es un error. Los organismos *no son copiados* durante la reproducción, ni son los causantes de su propia copia. Son formados de nueva planta de acuerdo con planos incorporados al ADN de los organismos de sus progenitores. Por ejemplo, un oso con el hocico deformado por un accidente tal vez cambie de estilo de vida a causa de su deformidad, lo cual podrá afectar para bien o para mal a su capacidad para sobrevivir y «reproducirse». Pero el oso con el hocico deforme no tiene ninguna posibilidad de ser *copiado.* Si tiene descendencia, ésta tendrá hocicos con la forma original. Sin embargo, si efectuamos alguna modificación en el correspondiente gen (sólo deberemos cambiar una molécula, si lo hacemos inmediatamente después de la concepción) del oso, su descendencia no sólo tendrá hocicos deformes, sino también copias del nuevo gen. Esto demuestra que la nueva forma de los hocicos de esos osos está causada por ese gen, y no por la

deformidad de ningún hocico. Así pues, la configuración del hocico del oso progenitor no condiciona la que tendrá el de su descendencia, mientras que la forma de sus genes contribuye, en cambio, a su propia replicación y a la conformación del hocico de sus descendientes.

Por lo tanto, un organismo es el entorno inmediato que copia a los auténticos replicantes: los genes de ese organismo. Tradicionalmente, el hocico de un oso y su madriguera habían sido clasificados respectivamente, como una entidad viva y una entidad no viva, pero esta distinción no está fundada en ninguna diferencia significativa. El papel del hocico del oso no es, en el fondo, distinto del de su madriguera. Ninguna de las dos cosas es un replicante, por más que aparezcan nuevas copias de ambas de modo constante. Tanto el hocico como la madriguera son meras partes del entorno que manipulan los genes del oso durante el proceso de hacerse replicar.

La noción de que la vida se basa en los genes —que considera a los organismos como parte del entorno de los genes— ha constituido, implícitamente, la base de la biología desde Darwin, pero no se le dio la debida importancia, al menos hasta los años sesenta, y no puede decirse que se comprendieran todas sus implicaciones hasta que Richard Dawkins publicó *El gen egoísta* (1976) y *El fenotipo expandido* (1982).

Vuelvo a la pregunta de si la vida es o no un fenómeno fundamental de la naturaleza. Ya he advertido contra la asunción reduccionista de que los fenómenos emergentes, como la vida, son necesariamente menos fundamentales que los físicos microscópicos. Sin embargo, todo lo que acabo de decir sobre la vida parece indicar que es un mero efecto secundario más, al final de una larga cadena de efectos secundarios. Y es que no son tan sólo las *predicciones* de la biología las que se sujetan, en principio, a las de la física, sino también sus explicaciones. Como he dicho, las grandes teorías explicativas de Darwin (en sus versiones modernas, como la de Dawkins) y de la bioquímica moderna *son* reductivas. Las moléculas vivas —genes— son, después de todo, moléculas, tan sujetas a las leyes de la física y la química como las no vivas. No contienen ninguna sustancia especial, ni presentan ningún atributo físico extraordinario. Simplemente, son, en determinados entornos, replicantes. La propiedad de ser replicante es altamente contextual, es decir, depende de intrincados detalles del entorno. Una entidad puede ser replicante en un entorno y no

en otro. Del mismo modo, la propiedad de estar adaptado a un nicho no depende de un simple e intrínseco atributo físico del que pueda hacer uso el replicante en un momento determinado, sino de los efectos que ese nicho pueda causar en él en el futuro en circunstancias hipotéticas (por ejemplo, variantes del entorno). Las propiedades contextúales e hipotéticas son, en esencia, derivativas, de modo que resulta difícil creer que un fenómeno caracterizado únicamente por ellas pueda ser fundamental en la naturaleza.

Por lo que se refiere al impacto físico de la vida, la conclusión es la misma: los efectos de la vida parecen despreciablemente pequeños. Por lo que sabemos, el planeta Tierra es el único lugar del universo en el que existe. El hecho es que no hemos encontrado pruebas de su existencia en ninguna otra parte, de modo que, aun en el supuesto de que estuviese ampliamente extendida, sus efectos serían demasiado pequeños para ser perceptibles para nosotros. Lo que vemos más allá de la Tierra es un universo activo, lleno de procesos diversos y poderosos, pero completamente inanimados. Las galaxias evolucionan. Las estrellas se condensan, brillan, flamean, estallan y se contraen. Partículas de alta energía y ondas electromagnéticas y gravitatorias fluyen en todas direcciones. Que la vida esté presente o no en esos titánicos procesos no parece tener mayor importancia. Todo indica que ninguno de ellos se vería afectado en lo más mínimo si *estuviera* presente. Si la Tierra se viese envuelta en una llamarada solar —lo que en sí mismo constituiría un insignificante acontecimiento astrofísico—, nuestra biosfera quedaría instantáneamente esterilizada. Esa catástrofe, sin embargo, tendría sobre el Sol el mismo efecto que una gota de lluvia sobre un volcán en erupción. Nuestra biosfera es, en términos de su masa, energía o cualquier otra dimensión astrofísica significativa, algo insignificante, incluso en comparación con la Tierra. Es, además, un tópico en astrofísica que el sistema solar consiste esencialmente en el Sol y Júpiter. Todo lo demás (incluyendo la Tierra) es considerado «impurezas». Más aún, el sistema solar es una porción nimia de nuestra galaxia, la Vía Láctea, inapreciable a su vez entre la multitud de galaxias conocidas en el universo. Parece, pues, que, como dice Stephen Hawking: «La raza humana no es más que escoria química sobre un planeta de dimensiones discretas, el cual órbita alrededor de una estrella nada notable en el borde de una galaxia que no se distingue en absoluto de los cientos de miles de millones de

galaxias restantes.»

La idea dominante en la actualidad es, pues, que la vida, lejos de ser el centro, ya sea desde el punto de vista geométrico, teorético o práctico, es de una insignificancia casi inconcebible. La biología es, en este contexto, una materia con un status parecido al de la geografía. Conocer la trama urbana de la ciudad de Oxford es importante para los que vivimos en ella, pero carece de todo interés para quienes nunca la visiten. De modo semejante, parece que la vida es una propiedad de una pequeña área, o quizás áreas, del universo, fundamental sólo para nosotros porque estamos vivos, pero sin ninguna trascendencia, teorética o práctica, en el vasto esquema de las cosas.

Pero, por sorprendente que parezca, esta apariencia es engañosa. Es una falsedad que los efectos físicos de la vida sean insignificantes, o que sea un mero derivativo desde el punto de vista teorético.

Como primer paso para explicarlo, permítaseme profundizar en mi anterior afirmación de que la vida es una forma de generación de realidad virtual. He empleado la palabra «ordenadores» para describir los mecanismos que ejecutan los programas de los genes en el interior de las células, pero esta terminología es, en realidad, un tanto ambigua. En comparación con los ordenadores de uso general que fabricamos artificialmente, hacen más en algunos aspectos y menos en otros. No resultaría fácil programarlos para procesar textos o factorizar grandes números. En cambio, ejercen un control interactivo extremadamente exacto sobre las respuestas de un complejo entorno (el organismo) ante todo lo que le pueda suceder. Este control, además, está dirigido a hacer que el entorno realice una retroacción sobre los genes de un modo específico (es decir, replicándolos), de manera que el efecto neto sobre los genes resulte tan independiente como sea posible de lo que esté sucediendo en el exterior. Esto es más que informática, es generación de realidad virtual.

La analogía con la tecnología humana de realidad virtual no es, sin embargo, perfecta. En primer lugar, si bien los genes se hallan inmersos, al igual que el usuario de la realidad virtual, en un entorno cuya constitución y comportamiento detallados están especificados por un programa (al que los propios genes dan forma material), los genes no *experimentan* ese entorno, puesto que carecen de sentidos y experiencias. En consecuencia, si bien

un organismo es una representación en realidad virtual especificada por sus genes, es una representación sin audiencia. En segundo lugar, el organismo no sólo está siendo representado, sino también fabricado. No se trata de «engañar» al gen para hacerle creer que hay un organismo, sino que éste se encuentra realmente allí.

No obstante, esas diferencias son insustanciales. Como he dicho, *toda* representación en realidad virtual fabrica físicamente el entorno representado. El interior de cualquier generador de realidad virtual en el acto de representar es, precisamente, un entorno físico y real, fabricado para tener las propiedades especificadas en el programa. Somos los usuarios quienes escogemos, en ocasiones, interpretarlo como un entorno distinto, que resulta que provoca las mismas sensaciones. En cuanto a la ausencia de usuario, consideremos explícitamente en qué consiste el papel del usuario de realidad virtual. En primer lugar, debe «golpear» al entorno para que éste lo «golpee» a su vez; en otras palabras, ha de interactuar con el entorno de un modo autónomo. En el caso biológico, ese papel es asumido por el hábitat exterior. En segundo lugar, el usuario debe proporcionar la *intención* que hay tras la representación. Es decir, no tiene sentido afirmar que una determinada situación constituye una representación en realidad virtual sin tener en cuenta el concepto de fidelidad en la representación. Como dije, la fidelidad de un entorno es la semejanza, percibida por el usuario, entre el entorno reproducido y el que se deseaba reproducir. Pero ¿qué significa la fidelidad si se trata de un entorno que nadie deseaba reproducir y nadie percibe? Significa el grado de adaptación de los genes a su nicho. Podemos inferir la «intención» de los genes de reproducir un entorno que los replique de acuerdo con la teoría de la evolución de Darwin. Los genes se extinguen si no plasman esa «intención» tan eficiente y resueltamente como otros genes que compiten con ellos.

Así pues, los procesos vitales y las representaciones en realidad virtual son, diferencias superficiales aparte, la misma clase de proceso. Ambos implican la incorporación física de teorías generales sobre un entorno. En ambos casos esas teorías son utilizadas para construir el entorno y controlar, de modo interactivo, no tan sólo su apariencia instantánea, sino su respuesta detallada ante estímulos generales.

Los genes dan forma material al conocimiento sobre sus

nichos. Todo lo que tiene importancia fundamental en el fenómeno de la vida depende de esa propiedad, y no de la replicación *per se.* Podemos, pues, llevar ahora la discusión más allá de los replicantes. En principio, podríamos imaginar una especie cuyos genes fuesen incapaces de replicarse, pero que, en cambio, estuviesen adaptados para mantener invariable su aspecto físico mediante automantenimiento continuo y protección ante las influencias externas. No es probable que una especie así evolucionase de manera espontánea, pero podría ser obtenida de modo artificial. Al igual que el grado de adaptación de un replicante se define como el grado en que contribuye causalmente a su propia replicación, podemos definir el grado de adaptación de esos genes no replicantes como el grado en que contribuyen a su propia supervivencia en una determinada forma. Imaginemos una especie cuyos genes fuesen estructuras grabadas en la superficie de un diamante en bruto. Un diamante en bruto, tenga la forma que tenga, puede sobrevivir casi eternamente en gran variedad de circunstancias, pero esa forma no está *adaptada* para la supervivencia, ya que un diamante con una forma distinta sobreviviría también en circunstancias similares. Pero si los genes codificados en un diamante de nuestra especie hipotética fueran los causantes de que ese organismo se comportase de una manera determinada que, por ejemplo, protegiese la superficie grabada del diamante de la corrosión de un entorno hostil, o la defendiese de otros organismos que tratasen de introducir información distinta, o de ladrones que lo quisieran tallar y pulir para convertirlo en una gema, contendría entonces genuinas adaptaciones para la supervivencia en dichos entornos. (Incidentalmente, una gema *posee* cierto grado de adaptación para la supervivencia en el actual entorno terrestre. Los humanos buscan los diamantes en bruto y transforman sus formas naturales en gemas, pero también buscan las gemas y respetan sus formas. Así pues, en ese entorno, la forma de una gema contribuye causalmente a su supervivencia.)

Si cesase la fabricación de esos organismos artificiales, no podría volver a incrementarse el número de casos de cada gen no replicante, pero tampoco decrecería mientras el conocimiento que contuviesen fuera suficiente para permitirles mantener su estrategia de supervivencia para el nicho ocupado. Con el tiempo, un cambio lo bastante importante en el hábitat, o el deterioro causado por accidentes, podría conllevar la eliminación de la

especie, pero, en caso contrario, ésta podría muy bien sobrevivir tanto como cualquier especie natural. Los genes de esas especies comparten las propiedades de los genes reales, excepto la replicación. En particular, incorporan el conocimiento necesario para «simular» sus organismos, es decir, reproducirlos, del mismo modo en que lo hacen los genes reales.

Es la *supervivencia del conocimiento*, y no por fuerza la de los genes o cualquier otro objeto físico, lo que constituye el denominador común entre genes replicantes y no replicantes. Estrictamente hablando, pues, lo que está o no adaptado a cierto nicho es, en realidad, un fragmento de conocimiento, más que un objeto físico. Si está adaptado, tendrá la propiedad de que, una vez establecido en ese nicho, tenderá a permanecer allí. En el caso de un replicador, el material físico que le da forma cambia sin cesar, y con cada replicación se forma una nueva copia compuesta de partes no replicantes. El conocimiento no replicante puede también materializarse sucesivamente en *distintas* formas físicas, como, por ejemplo, cuando se transfiere el sonido de un disco antiguo, primero a una cinta de audio y luego a un disco compacto. Podríamos imaginar otros organismos vivos artificiales basados en no replicantes que se comportasen del mismo modo, aprovechando cada oportunidad de recopiar el conocimiento de sus genes para incorporarlo al medio más seguro disponible. Quizás algún día nuestros descendientes lo hagan.

No creo que resulte justo denominar «inanimados» a estos organismos, pero la terminología no es lo más importante, sino el hecho de que, si bien la vida toda se basa en replicantes, de lo que realmente trata el fenómeno de la vida es del conocimiento. Podemos, pues, dar una definición de adaptación directamente en términos de conocimiento: *Una entidad está adaptada a su nicho cuando da forma material al conocimiento que causa que ese nicho mantenga la existencia de dicho conocimiento.* Nos estamos acercando a la razón por la que la vida es fundamental. La vida trata de la materialización física del conocimiento. En el capítulo 6 nos encontramos con una ley de física, el principio de Turing, que trata también de la materialización física del conocimiento. Nos dice que es posible dar forma material a las leyes de la física, en cuanto son de aplicación a cualquier entorno físicamente posible, a los programas de un generador de realidad virtual. Los genes son programas así. Aún más, cualesquiera otros programas de realidad virtual que existan, o puedan llegar a existir físicamente, son

efectos directos o indirectos de la vida. Los programas de realidad virtual que ejecutan nuestros ordenadores o nuestros cerebros, por ejemplo, son efectos indirectos de la vida humana. La vida es, pues, el medio, presumiblemente necesario, mediante el cual los efectos a los que se refiere el principio de Turing son aplicados en la naturaleza.

Si bien ello resulta alentador, no es suficiente para establecer que la vida sea un fenómeno fundamental. Eso se debe a que no he demostrado aún que el principio de Turing tenga el status de ley fundamental. Un escéptico podría argumentar que no lo tiene. Se trata de una ley sobre la materialización física del conocimiento, y el escéptico podría adoptar la postura de que el conocimiento no es algo fundamental, sino un limitado concepto antropocéntrico. En otras palabras, se trata de una de esas cosas importantes para nosotros por lo que somos —animales cuyo nicho ecológico depende de la creación y aplicación de conocimiento—, pero no significativas en un sentido absoluto. Para un oso koala, cuyo nicho ecológico depende de las hojas del eucalipto, dicho árbol es significativo. Para el simio *Homo sapiens*, manipulador de conocimiento, éste es significativo.

Pero el escéptico estaría en un error. El conocimiento no es significativo únicamente para el *Homo sapiens*, ni sólo en el planeta Tierra. Como he dicho, el que algo tenga o no un gran impacto físico no es determinante, pero sí relevante, para que sea importante en la naturaleza. Consideremos ahora los efectos astrofísicos del conocimiento.

La teoría de la evolución estelar —la estructura y el desarrollo de las estrellas—es uno de los grandes éxitos de la ciencia. (Fíjense en el conflicto terminológico: la palabra «evolución» significa en física desarrollo o, simplemente, movimiento, no variación ni selección.) Hace nada más un siglo, se desconocía incluso el origen de la energía solar. La física más avanzada de la época sólo proporcionaba la errónea conclusión de que, cualquiera que fuese su fuente de energía, el Sol no podía haber estado brillando durante más de cien millones de años. Resulta curioso que geólogos y paleontólogos ya sabían, por la evidencia fósil del desarrollo de la vida, que el Sol debía de haber estado brillando sobre la Tierra durante mil millones de años, por lo menos. Luego se descubrió la física nuclear y fue aplicada con gran detalle a la física del interior de las estrellas. Desde entonces, la teoría de la evolución estelar ha ido madurando. Comprendemos

ahora lo que provoca el brillo de las estrellas. Para la mayoría de clases de estrellas podemos predecir qué temperatura, color, luminosidad y diámetro presentarán en cada etapa de su historia, cuánto durarán, qué elementos formará la estrella mediante transmutaciones nucleares, etcétera. Esta teoría ha sido comprobada y confirmada mediante observaciones del Sol y otras estrellas.

Podemos, pues, utilizar esta teoría para predecir el futuro desarrollo del Sol. Nos dice que el Sol continuará brillando con gran estabilidad durante otros cinco mil millones de años, más o menos, luego se expandirá hasta alcanzar un diámetro un centenar de veces superior al actual y se transformará en una estrella gigante roja; luego empezará a centellear, su brillo aumentará repentinamente, con lo que pasará a ser una nova, se contraerá y se enfriará, hasta convertirse con el tiempo en una enana negra. Pero ¿le ocurrirá de veras todo esto al Sol? ¿Se han convertido ya *todas* las estrellas formadas unos cuantos miles de millones de años antes que el Sol, con idéntica masa y composición, en gigantes rojas, como predice la teoría? ¿Es posible que algunos procesos químicos en apariencia insignificantes, sucedidos en los planetas menores en órbita alrededor de esas estrellas, puedan haber alterado el curso de procesos nucleares y gravitatorios de masa y energía infinitamente superiores?

Si el Sol se convierte en una gigante roja, absorberá a la Tierra y la destruirá. Si algunos de nuestros descendientes, físicos o intelectuales, se encuentran aún sobre la Tierra por aquel entonces, quizás no deseen que eso ocurra. Harán, sin duda, todo lo que esté en su mano para impedirlo.

¿Es incuestionablemente cierto que no podrán hacer nada al respecto? Sin duda, nuestra tecnología actual es demasiado insignificante para ello, pero ni nuestra teoría de la evolución estelar, ni ninguna otra física conocida, nos da motivo para pensar que la tarea sea imposible. Bien al contrario, sabemos ya, en términos amplios, lo que ésta conllevaría (básicamente, extraer materia del Sol). Nos quedan varios miles de millones de años para perfeccionar nuestros planes ahora en ciernes y ponerlos en práctica. Si, llegado el caso, nuestros descendientes consiguieran salvarse de ese modo, ello significaría que nuestra teoría actual de la evolución estelar, aplicada en concreto a la estrella Sol, hace una predicción completamente equivocada. Y la razón para ello es que no toma en consideración el efecto de la vida en la evolución

estelar. Tiene en cuenta efectos físicos fundamentales, como las fuerzas nucleares y electromagnéticas, la gravedad, la presión hidrostática y la de la radiación, pero no la vida.

Parece razonable pensar que el conocimiento necesario para modificar el Sol de ese modo no puede evolucionar únicamente por selección natural, de modo que cualquier futura modificación del Sol tiene que depender de la vida *inteligente.* Ante esto, se puede argumentar que quizás sea una desmesurada e infundada asunción que la inteligencia pueda sobrevivir sobre la Tierra durante varios miles de millones de años y que, incluso en este caso, es también una asunción pensar que se posea entonces el conocimiento necesario para modificar el Sol. Una opinión corriente en la actualidad es que la vida inteligente sobre la Tierra corre el peligro de autodestruirse, bien como consecuencia de una guerra nuclear, bien por algún efecto secundario del avance tecnológico o la investigación científica. Muchas personas creen que si la vida inteligente sobrevive, sólo lo conseguirá suprimiendo el progreso tecnológico. Podrían, pues, temer que el desarrollo de la tecnología necesaria para la modificación de las estrellas fuera incompatible con la supervivencia durante el tiempo suficiente para utilizarla, y que, por consiguiente, la vida sobre la Tierra está predestinada, de un modo u otro, a no afectar a la evolución del Sol.

Estoy convencido de que semejante pesimismo es infundado y, como explicaré en el capítulo 14, tenemos todos los motivos para conjeturar que, con el tiempo, nuestros descendientes podrán modificar el Sol y hacer muchas más cosas. Sin duda, no podemos prever ni su tecnología ni sus deseos. Puede que escojan salvarse emigrando del sistema solar, o refrigerando la Tierra, o utilizando cualquier otro medio, inconcebible hoy para nosotros, qué no requiera intervenir en el Sol. Por otra parte, podrían decidir modificarlo mucho antes de lo necesario, para prevenir su entrada en la fase de gigante roja (dominando su energía más eficazmente, por ejemplo, o explotando para mejorar su calidad de vida los minerales que contiene en su interior). Sin embargo, lo que quiero dejar bien claro es que lo que ocurra no dependerá de nuestra capacidad de predicción, sino sólo de qué conocimiento tengan nuestros descendientes y cómo decidan aplicarlo. No podemos, pues, predecir el futuro del Sol sin tomar posición sobre el futuro de la vida sobre la Tierra, en general, y del conocimiento, en particular. El color que tenga el Sol dentro de diez mil millones

de años depende de la presión gravitatoria y radiactiva, la convección y la síntesis nuclear. No depende en absoluto de la geología de Yenus, la química de Júpiter o la disposición de los cráteres de la Luna. Pero depende de lo que le ocurra a la vida inteligente en el planeta Tierra. Depende de la política, la economía y el resultado de las guerras. Depende de lo que haga la gente, de qué decisiones tome, qué problemas resuelva, qué valores tenga y cómo enfoque el bienestar futuro de sus hijos.

No podemos evitar llegar a esta conclusión adoptando una teoría pesimista sobre las perspectivas de nuestra supervivencia. Semejante teoría no se desprende de las leyes de la física ni de ningún otro principio fundamental que conozcamos, y tan sólo puede ser justificada en términos humanos de alto nivel, tales como que «el conocimiento científico ha sobrepasado al conocimiento moral», o afirmaciones por el estilo. Al utilizar esa teoría como argumentación se reconoce de un modo implícito que las teorías sobre los asuntos humanos son necesarias para efectuar predicciones astrofísicas. Incluso en el supuesto de que la especie humana fracasase en sus esfuerzos por sobrevivir, ¿sería de aplicación la teoría pesimista a toda clase de inteligencia extraterrestre del universo? En caso contrario, si alguna clase de vida inteligente, en alguna galaxia, llegara a sobrevivir durante miles de millones de años, confirmaría que la vida es significativa en el desarrollo físico general del universo.

A lo largo y lo ancho de nuestra galaxia y el multiverso, la evolución estelar depende de si se ha desarrollado la vida inteligente y, en caso afirmativo, de dónde lo ha hecho, de los resultados de *sus* guerras y de cómo enfoca el futuro de *sus* hijos. Podemos, por ejemplo, predecir aproximadamente las proporciones de estrellas de distintos colores (o, más precisamente, de distintos tipos espectrales) que debería haber en nuestra galaxia. Para ello, tendremos que hacer algunas asunciones sobre cuánta vida inteligente hay allí y cuáles son sus manifestaciones (¡saber cuántas estrellas ha sido capaz de apagar ayudaría mucho!). Por ahora, nuestras observaciones dan como resultado que no existe vida inteligente fuera del sistema solar. Cuando nuestras teorías sobre la estructura de la galaxia de la que formamos parte estén más depuradas, podremos hacer predicciones más exactas, pero, de nuevo, únicamente sobre la base de asunciones sobre la distribución y el comportamiento de la inteligencia en ella. Si estas asunciones son incorrectas, haremos

predicciones erróneas sobre la distribución de los tipos espectrales. Pero, si son correctas y detectamos ciertas anomalías en la distribución de dichos tipos espectrales, ello podría ser una prueba de la presencia de inteligencia extraterrestre.

Los cosmólogos John Barrow y Frank Tipler han considerado los efectos astrofísicos que tendría la vida en caso de sobrevivir durante un largo período *después* que el Sol se hubiese convertido en una gigante roja. Han descubierto que, con el tiempo, causaría importantes cambios cualitativos en la estructura de nuestra galaxia y, más tarde, en la de todo el universo (volveré sobre estos resultados en el capítulo 14). Una vez más, cualquier teoría sobre la estructura del universo, en lo que no sean sus etapas iniciales, deberá tomar posición sobre lo que la vida hará o dejará de hacer en cada momento. No hay escapatoria posible: la historia futura del universo depende de la historia futura del conocimiento. Los astrólogos creían que los acontecimientos cósmicos influían en los asuntos humanos, mientras que la ciencia ha creído durante siglos que no se afectaban mutuamente. Ahora vemos que son los asuntos humanos los que influyen en los acontecimientos cósmicos.

Vale la pena reflexionar qué nos hizo subestimar el impacto físico de la vida. Fue el tener una visión demasiado restringida de las cosas, lo que no deja de ser irónico, puesto que los antiguos llegaron a un consenso que conseguía evitar nuestro error precisamente por tener una visión de las cosas aún más restringida. En el universo, *tal como lo vemos*, la vida no ha tenido ningún efecto astrofísico importante. Sin embargo, únicamente vemos el pasado, y, dentro de él, sólo con cierto detalle lo que se halla más próximo a nosotros espacialmente. Cuanto más lejos miramos en el universo, más nos adentramos en el pasado y menos detalle percibimos. Pero el pasado, incluso considerado en su totalidad —la historia del universo desde el Big Bang hasta nuestros días-, es tan sólo una mínima parte de la realidad física. Queda por transcurrir un espacio de tiempo muchísimo más largo entre el momento actual y el Big Crunch (si llega a suceder), por no mencionar los restantes universos. No podemos observarlo, pero cuando aplicamos nuestras mejores teorías al futuro de las estrellas, las galaxias y el universo, percibimos un amplio campo en el que la vida puede influir y, a largo plazo, llegar a dominar todo lo que ocurra, al igual que hace ahora en la biosfera de la Tierra.

El argumento convencional para justificar la insignificancia de la vida otorga demasiado peso a las cantidades observables y «voluminosas» como la dimensión, la masa y la energía. Para la visión restringida pasada y presente, ésas eran, y siguen siéndolo, buenas medidas de la importancia astrofísica, pero no existe razón alguna en el campo de la física para que continúe siendo así. Es más, la propia biosfera proporciona abundantes ejemplos contrarios a la aplicación general de dichas medidas de importancia. En el siglo III antes de Cristo, por ejemplo, la masa de la población humana era de unos diez millones de toneladas. Podríamos pensar que no es probable que los procesos físicos ocurridos en dicho siglo, que involucraban el movimiento de masas muchas veces superiores, pudiesen verse significativamente afectados por la presencia o la ausencia de seres humanos. Sin embargo, la gran muralla china, cuya masa es de unos trescientos millones de toneladas, fue construida en esa época. Mover millones de toneladas de rocas es una acción que los humanos realizan constantemente. En la actualidad, sólo son necesarias algunas docenas de humanos para remover una masa de un millón de toneladas en la explanación de una vía férrea o la perforación de un túnel. Ello se pone aún más de relieve si establecemos la comparación, todavía más lógica, entre la masa de la tierra y las rocas desplazadas y la de la pequeña parte del cerebro del emperador o del ingeniero de la que surgen las ideas (o memes) que harán que dicha masa sea removida. La especie humana, entendida como un todo (o, si se prefiere, su reserva de memes), dispone ya, probablemente, del conocimiento suficiente para destruir varios planetas, si su supervivencia dependiese de ello. Incluso la vida no inteligente ha transformado la superficie y la atmósfera de la Tierra de un modo que equivale a muchas veces su propia masa. Todo el oxígeno de nuestra atmósfera, por ejemplo, ha sido creado por las plantas y es, por consiguiente, un efecto secundario de la replicación de los genes (es decir, de las moléculas), que descendían de una sola molécula. La vida no obra sus efectos por ser mayor o tener más masa

o más energía que otros procesos físicos, sino por poseer más conocimiento. En términos de efecto final sobre los resultados de los procesos físicos, el conocimiento es, al menos, tan importante como cualquier otra cantidad física.

Ahora bien: ¿existe en el caso de la vida, como asumían en la antigüedad que debía ocurrir, una diferencia física básica entre los

objetos portadores de conocimiento y los no portadores, una diferencia que no depende ni de los entornos del objeto ni de sus efectos en el futuro remoto, sino únicamente de sus atributos físicos inmediatos? Por extraordinario que parezca, la hay. Para ver en qué consiste, debemos adoptar la perspectiva del multiverso.

Consideremos el ADN de un organismo vivo, por ejemplo, un oso, y supongamos que en alguna parte de uno de sus genes encontramos la secuencia TCGTCGTTTC. Esta cadena concreta de diez moléculas, tanto en el nicho especial formado por el resto del gen como en *su propio* nicho, es un replicante. Incorpora una pequeña, pero importante, cantidad de conocimiento. Supongamos ahora, para redondear la hipótesis, que encontramos un segmento de ADN basura (de no gen) en el ADN del oso que tiene también la secuencia TCGTCGTTTC. Esta secuencia, sin embargo, no merece ser considerada replicante, puesto que apenas contribuye a su replicación y no es portadora de conocimiento alguno. Se trata de una secuencia aleatoria. Tenemos, pues, dos objetos físicos, segmentos ambos de la misma cadena de ADN, uno de los cuales es portador de conocimiento, mientras que el otro es una secuencia aleatoria. Sin embargo, son *físicamente idénticos.* ¿Cómo puede el conocimiento ser una cantidad física fundamental si un objeto lo posee y otro que es idéntico a él no?

Puede serlo porque esos dos segmentos no son, en realidad, idénticos. Lo parecen tan sólo cuando son vistos desde algunos universos, como el nuestro. Veámoslos de nuevo, pero esta vez desde otros universos. Como no lo podemos hacer físicamente, lo haremos desde un punto de vista teórico.

Sabemos que el ADN de los organismos vivos está sujeto naturalmente a variaciones aleatorias —*mutaciones*— en la secuencia de las moléculas A, C, G y T. Según la teoría de la evolución, en los genes las adaptaciones y, por consiguiente, su propia subsistencia, dependen de que dichas mutaciones hayan tenido lugar. Gracias a ellas, las poblaciones de cualquier gen contienen un determinado grado de variaciones, y los individuos portadores de genes con niveles más altos de adaptación son los que tienden a tener más descendencia. La mayoría de las variaciones de un gen lo incapacitan para su replicación, puesto que la secuencia alterada ya no transmite las instrucciones a la célula para producir algo útil. Otras variaciones, simplemente, hacen menos probable la replicación, es decir, disminuyen el nicho

del gen.

Algunas, sin embargo, pueden ser portadoras de nuevas instrucciones que hagan *más* probable la replicación. De este modo tiene lugar la selección natural. Con cada generación de variaciones y replicaciones el grado de adaptación de los genes supervivientes tiende a incrementarse. Ahora bien, una mutación aleatoria, causada, por ejemplo, por el impacto de un rayo cósmico, origina variaciones no sólo dentro de la población del organismo en un universo, sino también entre universos. Un «rayo» cósmico es una partícula subatómica de alta i energía que, como el fotón emitido por la linterna, viaja en diferentes direcciones en universos distintos. Así pues, cuando un rayo cósmico impacta en una cadena de ADN y produce una mutación, algunas de sus contrapartidas en otros universos se quedarán sin sus correspondientes copias de esa cadena de ADN, mientras que otras impactarán en distintas posiciones y originarán mutaciones diferentes. De este modo, un solo rayo cósmico que incidiera sobre una sola molécula de ADN causaría, en general, la aparición de gran número de mutaciones diferentes en distintos universos.

Cuando consideramos la posible apariencia de un determinado objeto físico en otros universos, no debemos mirar tan lejos en el multiverso que nos resulte imposible identificar una contrapartida suya en el otro universo. Tomemos, por ejemplo, un segmento de ADN. En algunos universos, simplemente, no hay moléculas de ADN. En otros, existe el ADN, pero son universos tan distintos del nuestro, que no hay modo de identificar qué segmento del ADN que se encuentra en ellos corresponde al que estamos considerando en éste. Resulta ocioso preguntarse qué aspecto tendría nuestro segmento de ADN en semejantes universos, de modo que debemos considerar únicamente universos lo bastante similares al nuestro para que no surja esa disparidad. Podríamos considerar, por ejemplo, tan sólo aquellos universos en los que existan osos y en los que, además, se haya introducido una muestra del ADN de un oso en una máquina de análisis programada para imprimir diez letras que representen su estructura en una posición específica, relativa a ciertas referencias sobre una determinada cadena de ADN. La argumentación que sigue no se vería afectada aunque escogiésemos cualquier otro criterio razonable para identificar segmentos de ADN de características similares en universos cercanos.

Según cualquiera de esos criterios, el segmento del gen del

oso debe tener la misma secuencia en casi todos los universos próximos que en el nuestro. Ello se debe a que cabe presumir que está altamente adaptado, lo que significa que la mayor parte de sus variantes no conseguirían hacerse copiar en la mayor parte de las variantes de sus entornos, y no podrían, por tanto, aparecer en dicha localización en el ADN de un oso vivo. En cambio, cuando el segmento de ADN no portador de conocimiento experimenta prácticamente cualquier mutación, la versión mutante seguirá siendo capaz de ser copiada. A lo largo de generaciones de replicaciones habrán ocurrido múltiples mutaciones, la mayoría de las cuales no habrán afectado a la replicación. Por consiguiente, el segmento de ADN basura, a diferencia de su contrapartida en el gen, será completamente heterogéneo en distintos universos. Podría muy bien suceder que cada posible variación de su secuencia estuviera representada por un igual en el multiverso (que es lo que debería entenderse cuando se dice que su secuencia es estrictamente aleatoria).

Por consiguiente, la perspectiva del multiverso revela una estructura física adicional en el ADN del oso. En nuestro universo, contiene dos segmentos con la secuencia TCGTCGTTTC. Uno de ellos es parte de un gen, mientras que el otro no. En casi todos los universos cercanos, el primer segmento conserva la secuencia TCGTCGTTTC que presenta en el nuestro, pero el segundo varía enormemente de secuencia de un universo próximo a otro. Así pues, y desde la perspectiva del multiverso, los dos segmentos no son ni remotamente parecidos (figura 8.1).

Una vez más, tuvimos un punto de vista demasiado restrictivo y nos dejamos llevar a la falsa conclusión de que las entidades portadoras de conocimiento pueden ser físicamente idénticas a las no portadoras, lo cual, a su vez, proyectó dudas sobre el status fundamental del conocimiento. Pero ahora hemos vuelto prácticamente al punto de partida. Reconocemos que la antigua idea de que la materia viva tiene propiedades especiales se aproximaba mucho a la verdad: no es la materia viva, sino la materia *portadora de conocimiento*, la que es físicamente especial. Dentro de un universo parece irregular, pero en el multiverso

presenta una estructura tan regular como la de un cristal.

En definitiva, el conocimiento es una cantidad física fundamental, y el fenómeno de la vida también, aunque en medida ligeramente inferior.

Imaginemos que observamos una molécula del ADN de una célula de oso con un microscopio electrónico y tratamos de distinguir las secuencias de gen de las de no gen, así como de estimar el grado de adaptación de cada gen. En cualquier universo, considerado aisladamente, esa tarea resultaría imposible. La propiedad de ser un gen —es decir, de estar altamente adaptado— es abrumadoramente complicada para que sea posible detectarla en un universo aislado. Se trata de una propiedad emergente. Deberíamos hacer múltiples copias del ADN con sus variaciones, utilizar la ingeniería genética para obtener numerosos embriones de oso para cada variante del ADN, esperar a que los osos crecieran y vivieran en diversos entornos representativas del nicho ecológico de su especie y ver cuáles de ellos conseguían procrear.

Pero con un microscopio mágico, capaz de ver en otros universos (lo que, insisto, no es posible: utilizamos una teoría para imaginar —o representar— lo que sabemos que debe existir allí), la tarea resultaría sencilla. Como en la figura 8.1, los genes se distinguirían de los no genes, del mismo modo que los campos cultivados se distinguen de los bosques en una fotografía aérea, o los cristales que han precipitado se destacan en una solución. Serían regulares en múltiples universos próximos, mientras que los no genes, los segmentos de ADN basura, serían irregulares. En cuanto al grado de adaptación de un gen, éste resultaría casi tan fácil de determinar. Los genes mejor adaptados presentarían la misma estructura en un amplio abanico de universos: tendrían «cristales» más grandes, por así decirlo.

Viajemos ahora a otro planeta para tratar de encontrar las formas de vida autóctonas, si existen. Una vez más, ésta es una tarea de notoria dificultad. Deberíamos realizar complejos y sutiles experimentos, cuyos innumerables fracasos han sido tema de numerosos relatos de ciencia ficción. Pero si pudiéramos, simplemente, observar el planeta con un telescopio «multiversal», la vida y sus consecuencias serían evidentes al primer vistazo. Sólo habría que buscar estructuras complejas que pareciesen irregulares en cualquier otro universo, pero idénticas en muchos universos próximos. Si encontrásemos alguna, habríamos hallado

la evidencia de un conocimiento materializado físicamente. Donde hay conocimiento, debe haber vida, al menos en el pasado.

Comparemos ahora a un oso con la constelación de la Osa Mayor. El oso es, anatómicamente, muy similar en muchos universos próximos. No son tan sólo sus genes los que presentan esa propiedad, sino todo su cuerpo (si bien otros atributos de este último, como el peso, variarán mucho más que sus genes; ello se debe a que, por ejemplo, en los distintos universos el oso habrá tenido diferente fortuna en su búsqueda de alimento). Pero en el caso de la constelación de la Osa Mayor no se da esta regularidad de un universo a otro. La configuración de esa constelación es el resultado de las condiciones iniciales del gas galáctico del que se formaron las estrellas. Estas condiciones fueron aleatorias —muy diversas en distintos universos, a nivel microscópico— y el proceso de formación de las estrellas a partir del gas tuvo como consecuencia diversas inestabilidades, que aumentaron la escala de las variaciones. El resultado es que la conformación de estrellas que vemos en esa constelación existe tan sólo en un abanico muy limitado de universos. En la mayoría de las variantes próximas de nuestro universo también hay constelaciones en el cielo, pero tienen configuraciones diferentes en cada una de ellas.

Por último, contemplemos del mismo modo el universo que nos rodea. ¿Qué captará la atención de nuestra mirada mágicamente potente? En un universo aislado, las estructuras más impresionantes son las galaxias y los cúmulos de galaxias. Pero tales objetos no tienen una estructura discernible a lo largo y lo ancho del multiverso. Donde aparece una galaxia en un universo, habrá una miríada de galaxias, con geografías muy distintas, en el multiverso. Y ello es así en todos los aspectos. Los universos cercanos sólo se parecen en ciertas características macroscópicas aproximadas, como requieren las leyes de la física, que son de aplicación a todos ellos. Así pues, la mayoría de las estrellas serán aproximadamente esféricas en cualquier parte del multiverso, y la mayor parte de las galaxias serán elípticas o espirales. Pero nada puede extenderse profundamente por universos cada vez más lejanos sin que su estructura detallada cambie de modo irreconocible. Sólo son excepciones de esta regla los escasos lugares en que el conocimiento se ha materializado físicamente. En ellos los objetos se extienden por grandes cantidades de universos sin perder por ello las características que los hacen reconocibles. Quizás la Tierra sea el único lugar así de nuestro universo, de

momento. En cualquier caso, esos lugares se distinguen, en el sentido que he descrito, por ser aquellos en que se localizan los procesos —la vida y el pensamiento- que han creado las mayores estructuras distintivas que hay en el multiverso.

TERMINOLOGÍA

Replicante. Entidad que causa que determinados entornos la copien.

Gen. Replicante molecular. La vida sobre la Tierra se basa en genes constituidos por ADN (ARN en el caso de algunos virus).

Meme. Idea que se comporta como un replicante (por ejemplo, una teoría científica).

Nicho. Para un replicante, conjunto de entornos en los que podría replicarse. Para un organismo, conjunto de entornos y estilos de vida en los que podría vivir y reproducirse.

Adaptación. Un replicante está adaptado a un nicho cuando causa su propia replicación en dicho nicho. De modo más general, una entidad está adaptada a su nicho cuando da forma material al conocimiento que causa que ese nicho mantenga la existencia de dicho conocimiento.

SUMARIO

Desde los tiempos de Galileo, el progreso científico parece haber refutado la antigua idea de que la vida es un fenómeno fundamental de la naturaleza. Ha puesto de manifiesto la vasta escala del universo, en comparación con la biosfera terrestre. La biología moderna parece haber confirmado esa refutación al describir los procesos vitales en términos de replicantes moleculares, genes, cuyo comportamiento se rige por las mismas leyes de la física que son de aplicación a la materia inanimada. Sin embargo, la vida *está* asociada a un principio fundamental de la física —el principio de Turing—, puesto que es el medio por el que la realidad virtual fue originalmente realizada en la naturaleza. Asimismo y a pesar de las apariencias, la vida *es* un proceso importante en las escalas superiores tanto del tiempo como del

espacio. El futuro comportamiento de la vida determinará el futuro comportamiento de las estrellas y las galaxias. Y la estructura regular de mayor escala, *a lo largo y a lo ancho* de los universos, sólo existe allí donde se ha desarrollado la materia portadora de conocimiento, como cerebros o genes constituidos por segmentos de ADN.

Esta conexión directa entre la teoría de la evolución y la teoría cuántica es, a mi entender, una de las más impresionantes e inesperadas de las múltiples conexiones entre estas dos vías. Otra de ellas es la existencia de una teoría cuántica de la calculabilidad *que constituye la base de la teoría de la calculabilidad que conocemos. Dicha conexión es el tema del próximo capítulo*

9. LOS ORDENADORES CUÁNTICOS

Para un neófito en la materia, el término *cálculo cuántico* puede sonar como el nombre de una nueva tecnología, quizás el último grito en la notable sucesión que comprende el cálculo mecánico, el cálculo electrónico transistorizado, el cálculo mediante chips de silicio, etcétera. Por otra parte, es cierto que la tecnología informática actual se basa en procesos microscópicos cuantomecánicos. (Por supuesto, *todos* los procesos físicos son cuantomecánicos, pero aquí me refiero a aquellos para los que la física clásica —es decir, la física no cuántica— proporciona predicciones muy poco exactas.) Para que continúe la tendencia actual hacia un hardware cada vez más rápido y más compacto, la tecnología deberá ser cada vez más «cuantomecánica», en el sentido aludido, por la sencilla razón de que los fenómenos cuantomecánicos son los que predominan en todos los sistemas lo bastante pequeños. Si esto fuera todo, el cálculo cuántico mal podría figurar en ninguna explicación fundamental de la estructura de la realidad, ya que no habría nada sustancialmente nuevo en él. Los ordenadores actuales, cualquiera que sea la cantidad de procesos cuanto-mecánicos que utilicen, no son más que aplicaciones tecnológicas meramente distintas de la misma idea clásica, la de la máquina universal de Turing. Por esta razón el repertorio disponible de cálculos es, básicamente, el mismo para la totalidad de los ordenadores existentes, y difieren tan sólo en su velocidad, capacidad de memoria y periféricos de entrada y salida. Es decir, incluso el más modesto ordenador personal actual puede ser programado para resolver cualquier problema o reproducir cualquier entorno con tanta perfección como el más potente, a condición, simplemente, de que se le proporcione la suficiente memoria adicional, se le conceda el tiempo necesario y se le dote del hardware adecuado para mostrar los resultados.

El cálculo cuántico es algo más que una tecnología más rápida y miniaturizada para llevar a la práctica las máquinas de Turing. Un *ordenador cuántico* es una máquina que utiliza únicamente efectos cuantomecánicos, en especial la interferencia, para resolver clases de cálculos completamente nuevos y que serían imposibles, incluso en principio, para cualquier máquina de Turing y, por consiguiente, para cualquier ordenador clásico. El cálculo cuántico es, por lo tanto, ni más ni menos que una nueva

manera de dominar la naturaleza.

Permítaseme argumentar esta afirmación. Los primeros inventos para dominar la naturaleza fueron herramientas movidas por músculos humanos. Revolucionaron la situación de nuestros antepasados, pero adolecían de la limitación de requerir atención y esfuerzo humanos constantes durante cada momento de su utilización. La tecnología subsiguiente superó esta limitación. Los seres humanos consiguieron domesticar algunos animales y cultivar ciertas plantas, y dirigieron las adaptaciones biológicas de estos organismos para sus fines. Las cosechas crecían y los perros guardianes vigilaban mientras sus dueños dormían. Se introdujo una nueva clase de tecnología cuando los seres humanos fueron más allá de la mera explotación de las adaptaciones de que disponían (y de los fenómenos no biológicos de que disponían, tales como el fuego) y crearon adaptaciones completamente nuevas en el mundo, tales como cerámica, ladrillos, ruedas, artefactos metálicos y máquinas. Para ello tenían que reflexionar sobre las leyes naturales que gobiernan el mundo y comprenderlas, lo que incluía, como he dicho, no tan sólo sus aspectos superficiales, sino la estructura de la realidad en que se basaban. Siguieron miles de años de progreso en esta clase de tecnología, en la utilización de algunos de los *materiales, fuerzas y energías* de la física. En el siglo XX la *información* vino a sumarse a la lista cuando la invención de los ordenadores hizo posible el procesado de información compleja fuera del cerebro humano. El *cálculo cuántico*, actualmente en su primera infancia, es un paso más en esta progresión. Será la primera tecnología que permita desarrollar tareas útiles mediante la colaboración de universos paralelos. Un ordenador cuántico será capaz de distribuir los componentes de una tarea compleja entre gran número de universos paralelos y compartir después los resultados.

Ya he mencionado la trascendencia de la universalidad del cálculo. el hecho de que un único ordenador físicamente posible puede, si se le da el tiempo suficiente y tiene la memoria adecuada, realizar cualquier cálculo que pueda desarrollar cualquier otro ordenador físicamente posible. Las leyes de la física, tal como las conocemos, admiten la universalidad del cálculo. Sin embargo, para que sea útil o trascendente de algún modo en el esquema general de las cosas, la universalidad, tal como la he definido hasta ahora, no resulta suficiente. Significa tan sólo que el ordenador universal puede hacer, *con el tiempo*, lo mismo que

cualquier otro ordenador. En otras palabras, *si se le da el tiempo suficiente*, será universal. Pero ¿qué sucederá si no dispone del tiempo suficiente? Imaginemos un ordenador universal que tan sólo pudiese efectuar un paso de un programa de cálculo en toda la vida del universo. ¿Seguiría siendo su universalidad una profunda propiedad de la realidad? Seguramente, no. Para exponerlo de un modo más general, esta estrecha noción de la universalidad es criticable porque clasifica a una tarea como parte del repertorio de un ordenador sin tomar en consideración los recursos físicos que éste consumiría para realizarla. Así, por ejemplo, podría darse el caso de que un usuario de realidad virtual tuviera que permanecer en un estado de total suspensión de los sentidos durante miles de millones de años mientras el ordenador calculaba lo que debía representar a continuación. Ésta era la actitud adecuada al debatir los límites últimos de la realidad virtual, pero cuando lo que consideramos es la *utilidad* de la realidad virtual —o, lo que es más importante, el papel fundamental que tiene en la estructura de la realidad—, debemos ser más selectivos. La evolución jamás habría despegado si la tarea de representar determinadas propiedades de los primeros hábitats más simples no hubiese sido *tratable* (es decir, calculable en un tiempo aceptable) mediante la utilización como ordenadores de las moléculas disponibles. Del mismo modo, la ciencia y la tecnología nunca hubiesen echado a volar si el diseño de una herramienta de piedra hubiese necesitado de mil años de reflexión. Lo que era cierto al principio ha seguido siendo una condición absoluta en cada paso del progreso. La universalidad del cálculo no sería de mucha utilidad para los genes, por más conocimiento que contuviesen, si reproducir sus organismos fuera una tarea intratable en la que, por ejemplo, un ciclo reproductivo durase miles de millones de años.

Así pues, el hecho de que *existan* organismos complejos, y de que *haya habido* una sucesión de invenciones y de teorías científicas cada vez mejores (como la mecánica de Galileo, la mecánica de Newton, la mecánica de Einstein, la mecánica cuántica, etcétera), nos dice algo sobre la clase de universalidad de cálculo que existe en la realidad. Nos dice que las leyes de la física actuales son, por lo menos hasta ahora, susceptibles de ser superadas en todo momento por teorías que proporcionen mejores predicciones, y que la tarea de descubrir cada nueva teoría a partir de la anterior ha sido tratada por medio del cálculo dadas las leyes

y la tecnología ya conocidas. La estructura de la realidad debe, pues, estar, por así decirlo, *escalonada*, en aras de una fácil autoaccesibilidad. Del mismo modo, si consideramos la propia evolución como un cálculo, ello nos indica que deben de haber existido los suficientes organismos viables, codificados por el ADN, para permitir a los mejor adaptados ser calculados (es decir, desarrollarse) empleando los recursos proporcionados por sus predecesores menos adaptados. Podemos, pues, inferir que las leyes de la física, además de prescribir taxativamente su propia comprensibilidad mediante el principio de Turing, aseguran que los correspondientes procesos evolutivos, como la vida y el pensamiento, no requieran, para ocurrir realmente, un exceso de tiempo o de cualesquiera otros recursos.

Así pues, las leyes de la física no sólo permiten (o, como he argumentado, *exigen*) la existencia de vida y pensamiento, sino que prescriben que éstos sean, de la manera más adecuada, eficientes. Para expresar esta crucial propiedad de la realidad, los análisis modernos de la universalidad postulan habitualmente ordenadores que son universales en un sentido aún más completo del que requeriría, en apariencia, el principio de Turing. No sólo son posibles los generadores universales de realidad virtual, sino que es también posible construirlos de manera que no requieran cantidades de recursos impracticablemente grandes para reproducir aspectos simples de la realidad. A partir de ahora, me referiré a la universalidad en este sentido, excepto cuando especifique algo distinto.

Pero ¿con cuánta eficiencia, exactamente, pueden ser reproducidos determinados aspectos de la realidad? ¿Qué cálculos, en otras palabras, son factibles dentro de un tiempo y un presupuesto dados? Ésta es la cuestión básica de la teoría de la complejidad del cálculo, que, como he dicho, consiste en el estudio de los recursos requeridos para realizar tareas calculatorias dadas. La teoría de la complejidad aún no está lo suficientemente integrada en la física para ofrecer muchas respuestas cuantitativas. No obstante, ha abierto un notable camino en la definición de una distinción útil y práctica entre tareas calculatorias *tratables* e *intratables*. El enfoque general queda mejor ilustrado con un ejemplo. Consideremos la tarea de multiplicar dos números bastante grandes, por ejemplo 4.220.851 por 2.594.209.

Muchos de nosotros recordamos el método que aprendimos siendo niños para realizar multiplicaciones así. Consiste en

multiplicar cada dígito del multiplicando por todos los dígitos del multiplicador para obtener sumandos, que después se sumarán para conocer el resultado final, en nuestro caso 10.949.769.651.859. Algunos se resistirán a conceder que este fatigoso procedimiento haga la multiplicación «tratable» en cualquier sentido corriente de la palabra (de hecho, existen métodos más eficientes para multiplicar grandes números, pero éste nos proporciona un ejemplo bastante bueno); sin embargo, desde el punto de vista de la teoría de la complejidad, que considera operaciones complejísimas desarrolladas por ordenadores indiferentes al aburrimiento y que rara vez se equivocan, este método entra, sin duda, dentro de la categoría de «tratable».

Lo que cuenta para la «tratabilidad», según las definiciones al uso, no es el tiempo empleado en multiplicar dos números determinados, sino el hecho de que dicho tiempo no se incremente de modo desmesurado al aplicar el mismo método a números cada vez mayores. Por sorprendente que parezca, este modo más bien indirecto de definir la tratabilidad funciona muy bien en la práctica para muchas clases importantes de tareas calculatorias. Por lo que se refiere a la multiplicación, podemos comprobar fácilmente que el método estándar puede ser usado con números diez veces mayores, por ejemplo, con muy poco trabajo adicional. Supongamos, en aras de la argumentación, que cada multiplicación elemental de un dígito por otro dura en un determinado ordenador un milisegundo (incluyendo el tiempo necesario para sumar, llevar las decenas y demás operaciones complementarias que siguen a cada multiplicación elemental). Cuando multiplicamos los números de siete dígitos 4.220.851 y 2.594.209, cada uno de los siete dígitos de 4.220.851 debe ser multiplicado por cada uno de los siete dígitos de 2.594.209. El tiempo total requerido para la multiplicación (si las operaciones se desarrollan de modo secuencial) será de siete veces siete, o sea, 49 milisegundos. Para números aproximadamente diez veces mayores, que tendrían ocho dígitos cada uno, el tiempo requerido para su multiplicación sería de 64 milisegundos, lo cual representa un incremento, de tan sólo el 31 por ciento.

Queda claro, pues, que una amplia gama de números —que incluye sin duda todos aquellos susceptibles de ser obtenidos como medida de los valores de variables físicas— puede ser multiplicada en una pequeña fracción de segundo. Así pues, la multiplicación es

ciertamente tratable a todos los efectos dentro de la física (o, al menos, de la física actual). Por supuesto, fuera de la física puede presentarse la necesidad de multiplicar números mucho mayores. Los productos de números primos de 125 dígitos o más, por ejemplo, son de gran interés para los criptógrafos. Nuestra máquina hipotética podría multiplicar dos de esos números, lo que arrojaría un resultado de 250 dígitos, en poco más de una centésima de segundo. En un segundo, podría multiplicar dos números de mil dígitos. Los ordenadores reales disponibles en la actualidad pueden llegar incluso a mejorar esos tiempos. Tan sólo algunos investigadores en ramas esotéricas de la matemática pura están interesados en multiplicaciones tan incomprensiblemente vastas, si bien, como vemos, ni siquiera ellos tienen motivos para considerar la multiplicación como intratable.

En cambio, la *factorización* —básicamente, lo inverso de la multiplicación- parece mucho más difícil. Se empieza con un solo número como dato inicial —por ejemplo, 10.949.769.651.859— y se trata de encontrar dos factores, es decir, dos números más pequeños que, multiplicados entre sí, arrojen dicho resultado. Como acabamos de multiplicarlos, sabemos que la respuesta, en este caso, es 4.220.851 y

2.594.209. Dado que ambos son números primos, ésta es la única respuesta correcta. Pero, si no hubiésemos conocido estos números de antemano, ¿cómo los habríamos encontrado? Por más que busquen en los recuerdos de su infancia, no encontrarán ningún método fácil, porque no lo hay.

El método más obvio consiste en dividir el número inicial por todos los factores posibles, empezando con 2 y continuando con todos los números impares, hasta encontrar uno que lo divida exactamente. Uno de los factores, al menos (si existen, es decir, en el caso hipotético de que el número inicial no sea primo), no podrá ser superior a la raíz cuadrada del número inicial, lo que da una idea del tiempo que puede necesitar este método. En el ejemplo que estamos considerando, nuestro ordenador encontraría el más pequeño de los factores (2.594.209) en aproximadamente un segundo. Sin embargo, un número inicial diez veces mayor tendría una raíz cuadrada unas tres veces mayor, de modo que su factorización por este método requeriría el triple de tiempo. En otras palabras, añadir un dígito al número inicial *triplicarla* el tiempo de ejecución. Un dígito más volvería a triplicarlo, y así sucesivamente. El tiempo de ejecución aumentaría, pues, en

progresión geométrica, es decir, de modo exponencial, con cada aumento de un dígito en el número inicial. Factorizar un número con factores de veinticinco dígitos mediante este método mantendría ocupados a todos los ordenadores de la Tierra durante siglos.

El método se podría mejorar, pero *todos* los métodos de factoriza-ción actualmente en uso presentan la propiedad del incremento exponencial. El mayor número jamás factorizado —que, en realidad, lo fue «con trampa», pues se trata de un número cuyos factores fueron previamente seleccionados por unos matemáticos para desafiar a otros— tenía 129 dígitos. Se consiguió realizar la factorización gracias a un llamamiento vía Internet, que originó un esfuerzo de cooperación de miles de ordenadores. El científico informático Donald Knuth ha estimado que la factorización de un número de 250 dígitos, utilizando los métodos más eficientes hasta ahora conocidos, duraría más de un millón de años si dispusiera de una red compuesta por un millón de ordenadores. Tales cosas son difíciles de estimar, pero, aun en el caso de que Knuth fuera excesivamente pesimista, sólo hay que considerar números con algunos dígitos más para comprobar que la tarea resulta varias veces más ardua. Eso es lo que queremos decir al afirmar que la factorización de grandes números es intratable. Todo ello es bien distinto del caso de la multiplicación, en el que, como hemos visto, multiplicar dos números de 250 dígitos es coser y cantar para cualquier ordenador doméstico. Nadie es capaz de concebir siquiera cómo factorizar números de mil dígitos o de un millón.

O, al menos, nadie *podía* hasta hace poco.

En 1982 el físico Richard Feynman empezó a estudiar la simulación mediante ordenador de objetos cuantomecánicos. Su punto de partida fue algo ya conocido, pero cuya trascendencia no había sido apreciada: que la predicción del comportamiento de sistemas cuantomecánicos (o, en otras palabras, la representación de entornos cuantomecánicos en realidad virtual) es, en general, una tarea intratable. Una de las razones por las que la trascendencia de este hecho había pasado inadvertida era que nadie esperaba que la predicción por ordenador de fenómenos físicos interesantes resultase especialmente fácil. Consideremos, por ejemplo, la predicción meteorológica, o la de los terremotos. Si bien las correspondientes ecuaciones son bien conocidas, la dificultad de aplicarlas a situaciones reales es notoria. Este hecho

ha llamado la atención del gran público gracias a algunos libros y artículos de divulgación sobre el *caos* y el *efecto mariposa.* Estos efectos no tienen nada que ver con la intratabilidad a la que se refería

Feynman, por la sencilla razón de que ocurren tan sólo en física clásica; por tanto, no pueden darse en la realidad, que es cuantomecánica. No obstante, deseo aclarar algunos puntos sobre los movimientos clásicos «caóticos», aunque sólo sea para destacar las grandes diferencias entre la impredecibilidad clásica y la cuántica.

La teoría del caos trata de las limitaciones de la predecibilidad en la física clásica, y parte del hecho de que casi todos los sistemas clásicos son intrínsecamente inestables. La «inestabilidad» en cuestión no tiene nada que ver con ninguna tendencia al comportamiento violento o a la desintegración, sino que se refiere a una extrema sensibilidad a las condiciones iniciales. Supongamos que conocemos el estado actual de algún sistema físico, por ejemplo, el de un conjunto de bolas que ruedan sobre una mesa de billar. Si el sistema obedeciese a las leyes de la física clásica, cosa que hace en un grado bastante elevado, estaríamos en condiciones de determinar con exactitud su comportamiento futuro —podríamos predecir, por ejemplo, si una determinada bola entrará o no en un determinado agujero— a partir de las leyes del movimiento, del mismo modo que podemos predecir un eclipse o una conjunción de planetas a partir de ellas. Pero, en la práctica, nunca llegamos a poder medir con exactitud las posiciones y velocidades iniciales. La pregunta está servida: ¿conociendo estas condiciones iniciales con un grado aceptable de exactitud, podríamos predecir con igual aproximación el comportamiento futuro del sistema? La respuesta es que, por lo general, no. La diferencia entre la trayectoria predicha, que ha sido calculada a partir de datos ligeramente inexactos, y la real tiende a crecer de modo exponencial e irregular («caóticamente») con el tiempo, de forma que el estado inicial, imperfectamente conocido, acaba por no servir de guía para el comportamiento del sistema. Ello implica, por lo que se refiere a las predicciones realizadas mediante ordenador, que los movimientos de los planetas, epítome de la predecibilidad clásica, son sistemas clásicos atípicos. Para predecir lo que hará un sistema clásico típico, incluso tras un período de tiempo moderado, deberíamos poder establecer su estado inicial con un grado tal de exactitud,

que resulta imposible. De ahí la afirmación de que, en principio, el aleteo de una mariposa en un hemisferio podría causar un huracán en el otro. La impracticabilidad de las predicciones meteorológicas, y de todas aquellas que se les asemejan, se atribuye, pues, a la imposibilidad de tomar en consideración el aleteo de todas las mariposas del planeta.

Sin embargo, en realidad, los huracanes y las mariposas obedecen a la teoría cuántica y no a la mecánica clásica. La inestabilidad, que aumentaría con rapidez cualquier falsa especificación de un estado clásico inicial, simplemente, no es una característica de los sistemas « cuantomecánicos. En mecánica cuántica las pequeñas desviaciones de un estado inicial específico sólo tienden a causar pequeñas desviaciones del estado final predicho. Ahora bien, la predicción exacta también se ve dificultada, aunque por un efecto muy distinto.

Las leyes de la mecánica cuántica requieren que un objeto que se encuentre inicialmente en una posición dada (en todos los universos) se «difunda» en el sentido del multiverso. Por ejemplo, un fotón y sus contrapartidas en otros universos salen juntos del mismo punto del filamento de una lámpara, pero luego se mueven en billones de direcciones distintas. Cuando medimos después lo que ha sucedido, también nos «difundimos», ya que cada una de nuestras copias ve lo ocurrido en su particular universo. Si el objeto en cuestión es la atmósfera de la Tierra, el huracán tal vez haya sucedido en, digamos, un 30 por ciento de los universos y no en el restante 70 por ciento. Subjetivamente, percibimos esto como un resultado único, impredecible o «aleatorio», si bien desde la perspectiva del multiverso todos los resultados han ocurrido de modo real. Esta multiplicidad en universos paralelos es la verdadera razón de la impredecibilidad de la meteorología. Nuestra incapacidad para medir de manera adecuada las condiciones iniciales resulta por completo irrelevante. Aun en el, caso de conocerlas a la perfección, se mantendría la multiplicidad y, por consiguiente, la impredecibilidad de la evolución. Por otro lado, y en contraste con el caso clásico, un multiverso imaginario, con condiciones iniciales tan sólo ligeramente distintas, no se comportaría de modo muy diferente del multiverso real: quizás sufriría huracanes en el 30,000001 por ciento, y no en el 69,999399 por ciento restante.

No es, pues, el aleteo de las mariposas lo que provoca, en realidad, los huracanes, puesto que el fenómeno clásico del caos

depende de un determinismo perfecto, que no se da en ningún universo aislado. Consideremos un grupo de universos idénticos en un instante en el que, en todos ellos, ha aleteado una determinada mariposa. Consideremos un segundo grupo de universos que, en ese mismo instante, son idénticos a los del primero, excepto en qu? en ellos la mariposa no ha aleteado. Esperemos unas horas. La mecánica cuántica predice que, a menos que se den circunstancias excepcionales (como que alguien esté observando las alas de la mariposa y oprima el detonador de una bomba atómica si se pone a aletear), los dos grupos de universos, casi idénticos al principio, seguirán siéndolo al final. Sin embargo, interiormente, ambos grupos se habrán diferenciado sobremanera el uno del otro. Incluirán universos con huracanes y sin ellos, algunos en los que la mariposa habrá cambiado espontáneamente de especie y habrá reajustado todos sus átomos, y otros en los que el Sol habrá estallado a causa de que los átomos de su núcleo entraron por azar en reacción nuclear. Con todo, los dos grupos serán muy parecidos. En los universos en los que aleteó la mariposa y hubo huracanes, éstos eran, ciertamente, impredecibles, pero la mariposa no fue la causante de que se formaran, puesto que también hubo huracanes casi idénticos en universos donde todo era igual que en aquéllos, excepto que la mariposa no aleteó.

Quizás resulte conveniente hacer hincapié en la diferencia entre *impredecibilidad* e *intratabilidad.* La impredecibilidad no tiene relación alguna con los recursos calculatorios disponibles. Los sistemas clásicos son impredecibles (o lo serían, si existiesen), a causa de su sensibilidad a las condiciones iniciales. Los sistemas cuánticos carecen de sensibilidad, pero son impredecibles a causa de su distinto comportamiento en diferentes universos, por lo que parecen aleatorios en la mayoría de éstos. En ambos casos, ningún cálculo, por complejo que sea, podrá reducir la impredecibilidad. La intratabilidad, en cambio, es una consecuencia de los recursos calculatorios. Se refiere a una situación en la que podríamos formular la predicción si fuésemos capaces de realizar los cálculos adecuados, lo cual resulta imposible porque los recursos necesarios son impracticablemente grandes. Para distinguir con claridad los problemas de impredecibilidad de los de intratabilidad en mecánica cuántica debemos considerar sistemas cuánticos que sean, en principio, predecibles.

A menudo se presenta la teoría cuántica como si sólo pudiese

hacer predicciones probabilísticas. Por ejemplo, en el experimento con la clase de interferencia formada por barrera perforada y pantalla, descrito en el capítulo 2, observamos que el fotón puede llegar a cualquier punto de la parte «luminosa» de la configuración de luces y sombras. Pero conviene tener presente que en muchos otros experimentos la teoría cuántica predice un único y definitivo desenlace. En otras palabras, predice que el experimento acabará teniendo el mismo resultado en todos los universos —aunque éstos difieran en los estadios intermedios de aquél— y, además, predice ese resultado. En tales casos, estamos observando *fenómenos de interferencia no aleatoria.* Tales fenómenos pueden ser demostrados con un interferómetro. Se trata de un instrumento óptico que consiste básicamente en un juego de espejos, algunos convencionales (figura 9.1) y otros semiazogados, como los utilizados en trucos de magia y comisarías de policía (figura 9.2). Cuando un fotón incide sobre la superficie de un espejo semiazogado, en la mitad de los universos rebota, igual que si fuera un espejo convencional, mientras que en la otra mitad lo atraviesa como si no existiera.

El fotón entra en el interferómetro por la parte superior izquierda, como muestra la figura 9.3. En todos los universos en que se realice el experimento, el fotón y sus contrapartidas viajan hacia el interferómetro por el mismo camino. Esos universos son, por lo tanto, idénticos.

Pero tan pronto como el fotón incide sobre el espejo semiazogado los universos inicialmente idénticos empiezan a tener comportamientos diferentes. En la mitad de ellos el fotón cruza el espejo y sigue su camino por la parte superior del interferómetro. En cambio, en la mitad restante, el fotón rebota en el espejo y desciende hacia el extremo inferior izquierdo del aparato. En ambos grupos de universos las versiones del fotón rebotan después en los espejos normales situados, respectivamente, en los extremos superior derecho e inferior izquierdo del aparato, y llegan de forma simultánea al espejo semiazogado situado en el extremo inferior derecho, donde se interfieren mutuamente. Recordemos que hemos permitido la entrada de un único fotón en el aparato, por lo que sigue habiendo un solo fotón en cada universo. En todos los universos este fotón ha incidido ahora sobre el espejo del extremo inferior derecho. En la mitad de ellos lo ha hecho desde la izquierda, mientras que en el resto lo ha hecho desde arriba. Las versiones del fotón en estos dos grupos de universos se interfieren por completo. El resultado final dependerá de la geometría exacta de la situación; la figura 9.3 muestra el caso en que, en todos los universos, el fotón acaba encaminándose hacia la derecha a través del espejo; en ninguno de ellos es transmitido o reflejado hacia abajo. Todos los universos son, pues, idénticos al final del experimento, tal como lo eran al principio. Se diferenciaron y se interfirieron mutuamente sólo durante una pequeña fracción de segundo entre el principio y el final de la experiencia.

Este notable fenómeno de interferencia no aleatoria constituye una evidencia de la existencia del multiverso tan incuestionable como la extraña conformación de las sombras que hemos visto en el capítulo 2. Y es que el resultado que he descrito es incompatible con cualquiera de los dos caminos posibles que hubiese podido tomar una partícula de haber un solo universo. Si proyectamos un fotón hacia la derecha a lo largo del tramo inferior del interferòmetro, por ejemplo, *podría* pasar a través del espejo semiazogado igual que hace el fotón en el experimento de interferencia, pero también podría ser que no: a veces sería

desviado hacia abajo. Del mismo modo, un fotón proyectado hacia abajo a lo largo del tramo vertical derecho podría ser desviado hacia la derecha, como sucede en el experimento de interferencia, pero también podría seguir en línea recta. Así pues, cualquiera que sea el camino por el que introduzcamos un fotón dentro del aparato, su salida será aleatoria. Tan sólo cuando se produce la interferencia entre los dos caminos es predecible el resultado. En consecuencia, lo que se halla dentro del aparato justo antes del final del experimento de interferencia no puede ser un solo fotón que sigue un único camino (por ejemplo, un fotón que viaje por el tramo inferior). Tiene que haber algo más que le impide ser desviado hacia abajo. Por ende, tampoco puede haber un único fotón que descienda por el tramo de la derecha. Tiene que haber algo allí que le impide seguir en línea recta hacia abajo, como haría en ocasiones de hallarse solo. Al igual que con las sombras, podemos realizar experimentos adicionales que nos confirmarán que ese «algo más» tiene todas las propiedades de un fotón que viaja por el otro camino e interfiere con el fotón que vemos, pero con nada más de nuestro universo.

Puesto que únicamente hay dos clases de universos en ese experimento, el cálculo de lo que ocurrirá sólo requerirá, más o menos, el doble de tiempo que si la partícula obedeciese a las leyes clásicas; por ejemplo, si estuviésemos calculando el recorrido de una bola de billar. Un factor de dos difícilmente convertirá estos cálculos en intratables. Sin embargo, ya hemos visto que es fácil, hasta cierto punto, alcanzar una multiplicidad de grado muy superior. En los experimentos con sombras un único fotón atraviesa una barrera en la que hay algunos pequeños orificios y cae sobre una pantalla. Supongamos que la barrera tuviese mil orificios. Hay lugares de la pantalla donde el

fotón puede caer (cae, en algunos universos) y otros en los que no puede caer. Para calcular si un punto dado de la pantalla puede llegar o no a recibir jamás un fotón, deberemos calcular los efectos de mutua interferencia de mil versiones del fotón en universas paralelos. De modo específico, deberemos calcular mil trayectorias desde la barrera hasta el punto dado de la pantalla y calcular después los efectos de cada fotón sobre los demás, para determinar si podrán o no alcanzar el punto dado de la pantalla. Deberemos, pues, realizar, aproximadamente, mil veces más cálculos que si estuviésemos investigando si una partícula clásica incidirá o no en el punto en cuestión.

La complejidad de esta clase de cálculos nos demuestra que en un entorno cuantomecánico sucede mucho más de lo que se ve. Y, utilizando el criterio de realidad del doctor Johnson en términos de complejidad calculatoria, hemos visto que esta propia complejidad es el motivo fundamental por el que carece de sentido negar la existencia del resto del multiverso. Pero pueden aparecer multiplicidades muy superiores cuando dos o más partículas que interactúan están involucradas en un fenómeno de interferencia. Supongamos que cada una de un par de partículas que interactúan tiene acceso a un millar de trayectorias disponibles. El par puede, pues, estar en un millón de estados distintos en cualquier estadio intermedio del experimento, de modo que es posible que haya un millón de universos que difieran en lo que el par de partículas esté haciendo en un momento determinado. Si las partículas en juego fuesen tres, el número de universos distintos sería de mil millones, de un billón para cuatro, y así sucesivamente. Por consiguiente, el número de historias que hay que calcular si queremos predecir lo que sucederá en tales casos aumenta de modo exponencial con el número de partículas que interactúan. Por esta razón, calcular el comportamiento de un sistema cuántico típico constituye una tarea realmente intratable.

Ésta es la intratabilidad con la que trabajó Feynman. Como podemos ver, no tiene nada que ver con la impredecibilidad; bien al contrario, se manifiesta con la máxima claridad en fenómenos cuánticos altamente predecibles. Ello se debe a que en estos fenómenos se da el mismo resultado final en todos los universos, pero dicho resultado es el producto de la interferencia entre multitud de universos, que son diferentes durante el transcurso del experimento. Todo ello es, en principio, predecible a partir de la teoría cuántica y no demasiado sensible a las condiciones iniciales. Lo que hace difícil predecir que el resultado de tales experimentos será siempre el mismo es la

cantidad, extraordinariamente grande, de cálculos necesaria para ello.

La intratabilidad es, en principio, un impedimento mucho más grande para la universalidad de lo que podría llegar a ser nunca la im-predecibilidad. Como ya he dicho, una representación perfectamente fiel de una ruleta no tiene por qué —mejor dicho, no debe— dar la misma secuencia de números que la real. Del mismo modo, no podemos preparar de antemano una representación en realidad virtual del tiempo que hará mañana.

Pero sí podemos (o algún día podremos) realizar una representación del tiempo que, si bien no será idéntica a las condiciones climatológicas de un día concreto, será tan realista en su comportamiento que ningún usuario, por experto que sea, podrá distinguirla de la realidad. El mismo principio es de aplicación para todo entorno no sometido a los efectos de la interferencia cuántica (es decir, la mayoría de los entornos). Reproducir tales entornos en realidad virtual constituye una tarea calculatoria tratable. Sin embargo, todo indica que no es posible la representación práctica de entornos sometidos a los efectos de la interferencia cuántica. Sin realizar cálculos exponencialmente grandes, ¿cómo podremos estar seguros de que en tales casos nuestro entorno representado no hará cosas que el entorno real no haría absolutamente nunca a causa de algún fenómeno de interferencia?

Podría parecer lógico concluir que la realidad no presenta, después de todo, una genuina universalidad calculatoria, puesto que los fenómenos de interferencia no pueden ser representados de modo inútil. Feynman, sin embargo, sacó correctamente la conclusión contraria. En lugar de contemplar la intratabilidad de la tarea como un obstáculo, la tomó como una oportunidad. Si requiere tantos cálculos averiguar lo que sucederá en u-n experimento de interferencia, se dijo, el propio proceso de preparar el experimento y medir su resultado exigirá a su vez cálculos no menos complejos. Por consiguiente, razonó, después de todo, podría ser posible representar entornos cuánticos de manera eficiente, a condición de que el ordenador pudiera realizar experimentos en un objeto cuantomecánico real. El ordenador escogería qué mediciones tendría que hacer en un elemento auxiliar de hardware cuántico a medida que iba operando, e incorporaría los resultados de estas mediciones a sus cálculos.

El hardware cuántico auxiliar sería, de hecho, otro ordenador. Un interferómetro, por ejemplo, podría servir para ese fin y, al igual que cualquier otro objeto físico, ser concebido como un ordenador. Actualmente, lo denominaríamos *ordenador cuántico de uso específico.*

La «programaríamos» disponiendo los espejos según una determinada geometría, y luego proyectaríamos un solo fotón hacia el primer espejo. En un experimento de interferencia no aleatoria, el fotón siempre emergerá en una misma dirección, determinada por la disposición de los espejos, dirección que

podríamos interpretar como indicadora del resultado del cálculo. En un experimento más complejo, con varias partículas en juego, el cálculo podría fácilmente volverse intratable, como ya hemos explicado. Pero, puesto que podemos obtener el resultado con facilidad, con sólo realizar el experimento, no resulta, después de todo, realmente intratable. Debemos, pues, ser más cuidadosos con la terminología. Es evidente que hay tareas calculatorias que resultan intratables si intentamos realizarlas con cualquier ordenador existente, pero que serían tratables si utilizáramos objetos cuantomecánicos, como ordenadores de uso específico. (Nótese que el hecho de que los fenómenos cuánticos puedan ser utilizados para realizar cálculos de este modo, depende de que no estén sujetos al caos. Si el resultado de los cálculos fuese una función inhabitualmente sensible al estado inicial, «programar» el dispositivo poniéndolo en el estado inicial adecuado podría ser imposible.)

Utilizar un dispositivo auxiliar cuántico de este modo podría ser considerado hacer trampa, puesto que cualquier entorno resulta, obviamente, mucho más fácil de reproducir cuando se dispone de una copia extra sobre la que se pueden hacer mediciones durante la representación. No obstante, Feynman aventuró que quizás no sería necesario utilizar una copia literal del entorno que estuviera siendo reproducido, y que debía ser posible encontrar dispositivos auxiliares más fáciles de construir, pero cuyas propiedades de interferencia fueran análogas a las del entorno en cuestión. Un ordenador normal podría realizar entonces el resto de la representación trabajando a partir de la analogía entre el dispositivo auxiliar y el entorno final. Ello constituiría una tarea tratable, o, al menos, eso esperaba Feynman. Más aún, conjeturó —acertadamente- que todas las propiedades cuanto-mecánicas de cualquier entorno final podrían ser simuladas mediante dispositivos auxiliares de un tipo determinado, que especificó (en concreto, una disposición de átomos en movimiento que se interfieren mutuamente). Denominó *simulador universal cuántico* a cualquier dispositivo de esa clase.

Pero no se trataba de una única máquina, y, por lo tanto, no podía ser considerado un ordenador universal. Las interacciones a que se verían sometidos los átomos del simulador no podrían ser establecidas definitivamente, como en un ordenador universal, sino que deberían ser reajustadas para la simulación de cada entorno final. Y el principio de la universalidad reside en el hecho

de que sea posible programar una misma máquina, especificada de una vez por todas, para realizar cualquier cálculo posible o reproducir cualquier entorno físicamente posible. En 1985 demostré que la física cuántica permite la existencia de un *ordenador universal cuántico.* La prueba fue bastante sencilla. Todo lo que tuve que hacer fue imitar las construcciones de Turing, pero utilizando la teoría cuántica en lugar de la mecánica clásica, que él asumía implícitamente, para definir la física en la que se basa. Un ordenador universal cuántico podría llevar a cabo cualquier cálculo que cualquier otro ordenador cuántico (o cualquier ordenador basado en el principio de Turing) pudiese realizar, y podría representar cualquier entorno finito físicamente posible en realidad virtual. Es más, se ha demostrado ya que el tiempo y los demás recursos que necesitaría para hacerlo no aumentarían de modo exponencial al hacerlo la dimensión o el detalle del entorno que se deseara reproducir, por lo que los cálculos relevantes serían tratables, según los estándares de la teoría de la complejidad.

La teoría clásica de la calculabilidad, que constituyó la base in-cuestionada de la informática durante medio siglo, ha quedado obsoleta, excepto como aproximación esquemática, al igual que el resto de la física clásica. La teoría de la calculabilidad es ahora la teoría cuántica de la calculabilidad. Dije que Turing había utilizado implícitamente la «mecánica clásica» en su construcción. Si miramos hacia atrás, podemos ver ahora que la teoría clásica de la calculabilidad ni siquiera se adaptaba por completo a la física clásica y presagiaba en muchos aspectos la teoría cuántica. No es casualidad que la palabra *bit*, que representa la mínima cantidad posible de información que un ordenador puede manipular, signifique esencialmente lo mismo que *cuanto*, es decir, cantidad discreta. Las variables discretas (variables que no pueden tener una sucesión continua de valores) son ajenas a la física clásica. Por ejemplo, si una variable tiene sólo dos valores posibles, digamos 0 o 1, ¿cómo va de uno a otro valor? (Ya formulé esta pregunta en el capítulo 2.) En la física clásica debería saltar de manera discontinua, lo que es incompatible con el modo en que las fuerzas y los movimientos se comportan en la mecánica clásica. En la física cuántica, en cambio, no es necesario ningún cambio discontinuo, incluso en el caso de que todas las cantidades mensurables sean discretas. Funciona del modo que expondré a continuación.

Para empezar, imaginemos un mazo de cartas colocado encima de una mesa: el mazo sería el multiverso, y las cartas dispuestas las unas sobre las otras, los universos paralelos. (Este modelo, en el que los universos forman una secuencia, subestima enormemente la complejidad del multiverso, pero resulta suficiente para ilustrar lo que deseo aclarar aquí.) Alteremos ahora el modelo para tomar en consideración el hecho de que el multiverso no es un conjunto discreto de universos, sino un continuo, y que no todos los universos son distintos. De hecho, por cada universo presente también está presente un continuo de universos idénticos que comprende una proporción ciertamente minúscula, pero no nula, del multiverso. En nuestro modelo, esta proporción podría quedar expresada por el grueso de una carta, de modo que cada una de ellas representaría todos los grupos de universos de una determinada clase. No obstante, y a diferencia del grueso de una carta, la proporción de cada clase de universos cambia con el tiempo de acuerdo con las leyes cuantomecánicas del movimiento. En consecuencia, la proporción de universos con una determinada propiedad también cambia, y lo hace continuamente. En el caso de una variable discreta que cambie de 0 a 1, supongamos que la variable tiene el valor 0 en todos los universos antes de que se inicie el cambio y que, después de éste, tiene el valor 1 en todos ellos. Durante el cambio, la proporción de universos en los que el valor es 0 va descendiendo de manera uniforme del ciento por ciento hasta cero, mientras que la proporción en la que el valor es 1 aumenta de modo paralelo desde cero hasta el ciento por ciento. La figura 9.4 muestra un cambio así visto desde el multiverso.

Parece desprenderse de la figura 9.4 que si bien la transición de 0 a 1 es objetivamente continua desde la perspectiva del multiverso, permanece subjetivamente discontinua desde el punto de vista de cualquier universo individual (que podríamos representar, por ejemplo, con una línea horizontal en la mitad de la parte superior de la figura 9.4). No obstante, se trata tan sólo

de una limitación del diagrama, no es una característica real de lo que sucede. Si bien el diagrama parece sugerir que en cada instante hay un universo que «acaba de cambiar» de 0 a 1, puesto que ha «cruzado la frontera», la realidad es distinta. No podría ser de otro modo, puesto que el universo en cuestión es exactamente idéntico a todos aquellos en que el valor del bit es 1 en aquel momento, de modo que si los habitantes de uno de esos universos estuvieran experimentando un cambio discontinuo, también lo experimentarían los habitantes de todos los demás universos. Por consiguiente, ninguno de ellos puede tener esta experiencia. Nótese también que, como explicaré en el capítulo 11, la idea de que algo *se mueva* por el diagrama como en la figura 9.4, en la que ya está representado el tiempo, es, simplemente, una equivocación. En cada instante el bit tiene valor 1 en una determinada proporción de universos y 0 en otra. Todos esos universos, en todos esos tiempos, están ya representados en la figura 9.4 y no se mueven en absoluto.

Otra de las maneras en las que la física cuántica está implícita en la informática clásica consiste es que todas las aplicaciones prácticas de los ordenadores de la clase de Turing dependen de cosas como la materia sólida o los materiales magnetizados, que no podrían existir en ausencia de efectos cuantomecánicos. Por ejemplo, todo cuerpo sólido consiste en una disposición de átomos, compuestos a su vez por partículas con carga eléctrica (electrones y, en el núcleo, protones). Pero, según el caos clásico, ninguna disposición de partículas con carga eléctrica podría permanecer estable de acuerdo con las leyes clásicas del movimiento. Las partículas con carga positiva o negativa, simplemente, cambiarían de posición para colisionar entre sí, y la estructura se desintegraría. Tan sólo la intensa interferencia cuántica entre las diversas trayectorias seguidas por las partículas cargadas de universos paralelos impide semejantes catástrofes y hace posible la materia sólida.

La construcción de un ordenador universal cuántico está muy lejos de las posibilidades de la tecnología actual. Como he dicho, la detección de un fenómeno de interferencia requiere siempre que exista la adecuada interacción entre *todas* las variables que han sido distintas en los universos que contribuyen a la interferencia. Por consiguiente, cuantas más partículas interactuantes intervengan en el proceso, más difícil tenderá a ser el diseño de la interacción que ponga de manifiesto la interferencia, es decir, el

resultado del cálculo. Entre las múltiples dificultades técnicas que implica trabajar al nivel de un solo átomo o un solo electrón, una de las principales es la de prevenir que el entorno se vea afectado porque los diferentes subcálculos se interfieran a su vez. Y es que si un grupo de átomos que experimenta un fenómeno de interferencia afecta de manera diferencial a otros átomos del entorno, la interferencia ya no podrá ser detectada únicamente mediante mediciones del grupo original y, en consecuencia, éste no realizará ya ningún cálculo cuántico de utilidad. Es lo que se conoce como *discoherencia*. Debo añadir que este problema se presenta a menudo erróneamente del modo inverso: se nos dice que «la interferencia cuántica es un proceso muy delicado, que debe ser protegido de toda influencia exterior». Es un error. Las influencias exteriores pueden provocar imperfecciones menores, pero es el efecto del cálculo cuántico sobre el mundo exterior lo que causa la discoherencia.

De ahí los esfuerzos por diseñar sistemas submicroscópicos en los que las variables portadoras de información puedan interactuar entre sí, pero afectando lo menos posible a su entorno. Otra simplificación novedosa, exclusiva de la teoría cuántica de la calculabilidad, disminuye parcialmente las dificultades provocadas por la discoherencia: resulta que, a diferencia de lo que ocurre en la calculabilidad clásica, cuando es necesario diseñar elementos específicos de la lógica clásica, tales como Y, O y NO, la forma precisa de las interacciones carece prácticamente de importancia en el caso cuántico. Virtualmente, cualquier sistema a escala atómica de bits interactuantes, a condición de que no fuera discoherente, serviría para realizar cálculos cuánticos útiles.

Se conocen fenómenos de interferencia que involucran vastos números de partículas, tales como la superconductividad o la superfluidez, pero no parece que ninguno de ellos pueda ser utilizado para realizar cálculos interesantes. En el momento de escribir este texto, en laboratorio sólo se pueden realizar con cierta facilidad cálculos cuánticos de un bit. Los investigadores, sin embargo, confían en que puedan conseguirse *umbrales cuánticos* (el equivalente cuántico de los elementos de la lógica clásica) de dos o más bits en los próximos años. Se trata de los componentes básicos de los ordenadores cuánticos. Algunos físicos, encabezados por Rolf Landauer, de IBM Research, se muestran pesimistas sobre las posibilidades de ulteriores avances. Creen que la discoherencia no podrá jamás ser reducida más allá de un nivel que tan sólo

permitirá la ejecución de unos pocos pasos cuantocalculatorios consecutivos. La mayoría de los investigadores en este campo se muestran más optimistas, si bien ello quizás se deba a que tan sólo investigadores optimistas pueden escoger un campo como éste. Se han construido ya algunos ordenadores cuánticos de uso específico (véase más adelante) y, en mi opinión, los seguirán otros más complejos en cuestión de años, más que de décadas. Por lo que se refiere al ordenador cuántico universal, confío en que su construcción sea también únicamente cuestión de tiempo, si bien no me atrevo a vaticinar si éste serán décadas o siglos.

El hecho de que el repertorio del ordenador cuántico universal contenga entornos cuya reproducción es clásicamente intratable, implica que también deben de haberse convertido en tratables nuevas clases de cálculos puramente matemáticos, puesto que las leyes de la física están, como dijo Galileo, expresadas en lenguaje matemático, y reproducir un entorno equivale a evaluar determinadas funciones matemáticas. Y, ciertamente, se ha hallado que muchas tareas matemáticas que resultaban intratables por todos los métodos clásicos podrían ser realizadas con eficacia mediante el cálculo cuántico. La más espectacular es la de factorizar grandes números. El método, conocido como *algoritmo de Shor*, fue descubierto por Peter Shor, de los Laboratorios Bell. (Mientras este libro estaba en galeradas, han sido descubiertos otros algoritmos cuánticos, como el *algoritmo de Grover*, para la búsqueda rápida en grandes listados.)

El algoritmo de Shor es extraordinariamente simple y requiere bastante menos hardware del que exigiría un ordenador universal cuántico. Parece probable, pues, que se pueda construir una máquina cuántica de factorización mucho antes de que toda la gama de cálculos cuánticos sea tecnológicamente factible. Ésta es una perspectiva de gran importancia para la *criptografía* (la ciencia que estudia los métodos para enviar y verificar la información de un modo seguro). Las redes de comunicación reales pueden ser globales y tener grandes y cambiantes números de participantes, por lo que las pautas de comunicación pueden resultar imprevisibles. Resulta impracticable que cada pareja de interlocutores intercambie, físicamente y por adelantado, claves criptográficas secretas que les permitan tener la seguridad de comunicarse sin temor a los intrusos. La c*riptografía de clave pública* ha sido introducida con este objeto, y se basa en métodos de codificación y descodificación cuyos usuarios no necesitan una

información secreta para comunicarse sin peligro. El método más seguro conocido de criptografía de clave pública se basa en la intratabilidad del problema de factorizar grandes números. Se conoce como cripto-sistema RSA, en honor de Ronald Rivest, Adi Shamir y Leonard Adelman, quienes lo idearon en 1978, y se basa en un procedimiento matemático por el que se codifica el mensaje utilizando un número grande (por ejemplo, de 250 dígitos) como clave. El destinatario podrá hacer pública tranquilamente esta clave, puesto que cualquier mensaje codificado con ella sólo podrá ser descodificado mediante el conocimiento de los factores del número. Podemos así escoger dos números primos de 125 dígitos y mantenerlos secretos, pero multiplicarlos y hacer público el resultado de 250 dígitos. Cualquiera podrá enviar un mensaje con esta clave, pero sólo quien conozca los factores secretos podrá descifrarlo.

Como he dicho, no existen perspectivas prácticas de factorizar números de 250 dígitos por medios clásicos, pero una máquina cuántica de factorización que utilizara el algoritmo de Shor sería capaz de hacerlo con tan sólo unos pocos miles de operaciones aritméticas, que podrían muy bien durar unos minutos. De este modo, cualquiera que tuviese acceso a una máquina así podría fácilmente leer cualquier mensaje interceptado que hubiese sido codificado utilizando el criptosistema RSA.

De poco les serviría a los criptógrafos seleccionar números grandes como claves, puesto que los recursos requeridos por el algoritmo de Shor sólo aumentan ligeramente con el tamaño del número que se ha de factorizar. En la teoría cuántica de la calculabilidad, la factorización es una tarea muy tratable. Se cree que, en presencia de un determinado nivel de discoherencia, habría de nuevo un límite práctico para el tamaño máximo del número que se podría factorizar, pero no se conoce un límite mínimo para el grado de discoherencia que pueda ser alcanzable tecnológicamente. Podemos, pues, concluir que, en el futuro, el criptosistema RSA, cualquiera que sea la longitud de su clave, resultará inseguro. En cierto sentido, ello hace que sea inseguro incluso en la actualidad, puesto que cualquier persona u organización que intercepte un mensaje en criptosistema RSA y espere hasta que le sea posible adquirir una máquina cuántica de factorización con una discoherencia lo suficientemente baja, podrá descifrarlo. Esto quizás tarde siglos en ocurrir, o quizás décadas, o quizás menos, ¿quién sabe? Pero la verosimilitud de que sea un

tiempo más bien largo es lo único que queda de la seguridad antaño total del criptosistema RSA.

Cuando una máquina cuántica de factorización está trabajando Con un número de 250 dígitos, la cantidad de universos que se interfieren será del orden de 10^{500}. Este número abrumadoramente grande es la razón por la que el algoritmo de Shor hace tratable la factori-zación. Ya dije que el algoritmo requiere tan sólo unos cuantos miles de operaciones aritméticas. Con ello me refería, por supuesto, a unos cuantos miles de operaciones aritméticas para cada universo que contribuya a la respuesta. Todos estos cálculos tienen lugar en paralelo en los distintos universos y comparten sus resultados mediante la interferencia.

Quizás se estén preguntando cómo podemos persuadir a nuestras contrapartidas en unos 10^{500} universos para que se pongan a trabajar en nuestra tarea de factorización. ¿No tendrán ya sus propios planes para la utilización de sus ordenadores? Afortunadamente, no será necesaria ninguna persuasión. El algoritmo de Shor opera al principio tan sólo en un conjunto de universos *idénticos* entre sí, y sólo causa su diferenciación dentro de los confines de la máquina de factorización. Así pues, nosotros, que somos quienes especificamos el número que se ha de factorizar y quienes esperamos que se calcule la respuesta, somos idénticos en todos los universos. Hay, sin duda, muchos otros universos en los que hemos programado diferentes números o en los que jamás hemos construido la máquina de factorizar. Pero esos universos difieren del nuestro en demasiadas variables —o, más precisamente, en variables que no están hechas para interactuar del modo adecuado con la programación del algoritmo de Shor— y, en consecuencia, no interfieren con el nuestro.

El argumento del capítulo 2, aplicado a *cualquier* fenómeno de interferencia, rebate la idea clásica de que existe un único universo. Lógicamente, la posibilidad de cálculos cuánticos complejos no añade nada nuevo a un caso de por sí ya incuestionable, pero lo que sí le proporciona es impacto psicológico. Gracias al algoritmo de Shor, el argumento se ha visto considerablemente fortalecido. A aquellos que aún se aferran a la idea de un solo universo, les planteo la siguiente cuestión: *explíquenme cómo funciona el algoritmo de Shor*. No me refiero a predecir, simplemente, que funcionará, que no es más que una mera cuestión de resolver algunas ecuaciones incontrovertibles.

Me refiero a proporcionar una explicación. Mientras el algoritmo de Shor factorizaba un número utilizando unas 10^{500} veces los recursos calculatorios cuya presencia es evidente, ¿dónde estaba el número fac-torizado? Hay tan sólo unos 10^{80} átomos en todo el universo visible, cantidad realmente minúscula en comparación con 10^{500}, así que, si el universo visible constituyese toda la extensión de la realidad física, ésta no podría ni remotamente contener los recursos necesarios para factorizar un número tan grande. ¿Quién lo ha factorizado, entonces? ¿Cómo, y dónde, tuvo lugar el cálculo?

He considerado clases tradicionales de tareas matemáticas que los ordenadores cuánticos estarían en condiciones de desarrollar con mayor rapidez que las máquinas actuales. Pero existe una clase adicional de nuevas tareas, accesible a los ordenadores cuánticos, que ningún ordenador clásico podría realizar jamás. Por una extraña coincidencia, una de las primeras en descubrirse de esas tareas concierne también a la criptografía de clave pública. En esta ocasión no se trata de descifrar un sistema existente, sino de aplicar un sistema nuevo y absolutamente seguro de *criptografía cuántica.* En 1989, en la oficina del teórico Charles Bennett, de IBM Research, fue construido el primer ordenador cuántico operativo. Se trataba de un ordenador cuántico de uso específico, consistente en un par de dispositivos criptográficos cuánticos diseñados por el propio Bennett y por Gilíes Brassard, de la Universidad de Montreal. Era la primera máquina capaz de llevar a cabo cálculos no triviales que ninguna máquina de Turing podría realizar.

En el criptosistema cuántico de Bennett y Brassard los mensajes son codificados en los estados de fotones individuales emitidos por un láser. Si bien son necesarios muchos fotones para transmitir un mensaje (uno por cada bit, más muchos otros malogrados en múltiples interferencias), las máquinas pueden ser construidas con la tecnología existente, puesto que sólo necesitan realizar sus cálculos cuánticos con un fotón a la vez. La seguridad del sistema no se basa ya en la intratabilidad, bien clásica o cuántica, sino directamente en las propiedades de la interferencia cuántica, y esto es lo que le otorga una absoluta seguridad, inalcanzable con la física clásica. Por muchos cálculos que realizara, con cualquier clase de ordenador, incluso durante billones de años, un intruso que hubiese interceptado un mensaje cuantocodificado, no podría descifrarlo, puesto que cuando la

comunicación se establece a través de un medio que presenta interferencia, *toda intrusión será detectada.* Según la física clásica, nada puede impedir a un intruso con acceso físico a un medio de comunicación, como una línea de teléfono, instalar un dispositivo pasivo de escucha. Pero, como he explicado, cuando se efectúa una medición mediante un sistema cuántico, se alteran automáticamente sus subsiguientes propiedades de interferencia. El plan de comunicación del sistema se basa en este efecto. Las partes que establecen la comunicación activan una serie de experimentos de interferencia y los coordinan en un canal de comunicación que por lo demás es público. Sólo si la interferencia supera la prueba de que no ha habido intrusión pasarán a la segunda etapa del plan, consistente en utilizar parte de la información transmitida como clave criptográfica. En el peor de los casos, un intruso persistente podría impedir que tuviese lugar la comunicación (cosa mucho más fácil de conseguir, por supuesto, cortando la línea telefónica), pero, por lo que concierne a leer el mensaje, tan sólo lo podrá hacer el destinatario, y ello lo garantizan las leyes de la física.

Como la criptografía cuántica se basa en la manipulación de fotones individuales, presenta una importante limitación. Los fotones que transportan el mensaje (uno por bit) deben llegar intactos del emisor al receptor. Pero todos los sistemas de transmisión conllevan pérdidas, y, si éstas son demasiado importantes, el mensaje no llegará a su destino. La instalación de repetidores (remedio para este problema en los sistemas de comunicación actuales) comprometería la seguridad, puesto que el intruso podría, sin ser detectado, interceptar los mensajes en algún repetidor. Los mejores sistemas cuantocriptográficos existentes utilizan fibra óptica y tienen un alcance de unos diez kilómetros. Eso bastaría para dotar al distrito financiero de una ciudad, por ejemplo, de un sistema absolutamente seguro de comunicación interna. Puede que no esté lejos la comercialización de sistemas así, pero, para resolver el problema de la criptografía de clave pública en general —por ejemplo, para comunicaciones globales—, serán necesarios nuevos avances en la criptografía cuántica.

La investigación experimental y teorética en el campo del cálculo cuántico se está acelerando en todo el mundo. Han sido propuestas nuevas tecnologías, aún más prometedoras, para la construcción de ordenadores cuánticos, y continuamente se

descubren y analizan nuevos tipos de cálculo cuántico que presentan diversas ventajas sobre el cálculo clásico. Todos estos avances me parecen muy prometedores, y creo que algunos de ellos tendrán frutos tecnológicos. Pero, en lo que concierne a este libro, eso no es más que una cuestión accesoria. Desde una postura que considera los principios fundamentales, importa poco lo útil que el cálculo cuántico pueda llegar a ser, o si el primer ordenador cuántico universal se construirá la semana que viene, dentro de varios siglos o nunca. La teoría cuántica de la calculabilidad debe, en cualquier caso, ser parte integrante de la concepción del mundo de cualquiera que busque una comprensión básica de la realidad. Podemos descubrir (y, de hecho, lo estamos descubriendo), estudiándolos teoréticamente, lo que los ordenadores cuánticos tienen que decirnos sobre las conexiones entre las leyes de la física, la universalidad y diversas vías a primera vista no relacionadas de explicación de la estructura de la realidad.

TERMINOLOGÍA

Cálculo cuántico. Cálculo que requiere procesos cuantomecánicos, en particular, el de interferencia. En otras palabras, cálculo realizado mediante la colaboración entre universos paralelos.

Cálculo exponencial. Cálculo cuyas necesidades de recursos (como el tiempo) se incrementan de acuerdo con un factor aproximada-$_{r}$ mente constante cada vez que se añade un dígito al número inicial.

Tratable/Intratable (regla práctica aproximada). Una tarea de cálculo se considera tratable cuando los recursos necesarios para realizarla no aumentan de modo exponencial al añadir dígitos al número inicial.

Caos. Inestabilidad del movimiento en la mayoría de los sistemas clásicos. Una ligera diferencia entre dos estados iniciales origina desviaciones exponencialmente crecientes entre las dos trayectorias resultantes. Pero la realidad obedece a las leyes de la física cuántica, no de la clásica. Normalmente, la impredecibilidad provocada por el caos es absorbida por la indeterminación causada por el proceso mediante el cual universos idénticos se vuelven

diferentes.

Ordenador cuántico universal. Ordenador que podría llevar a cabo todo cálculo que cualquier otro ordenador cuántico fuera capaz de realizar, así como reproducir en realidad virtual cualquier entorno finito físicamente posible.

Criptografía cuántica. Cualquier forma de criptografía que pueda ser realizada por los ordenadores cuánticos, pero no por los clásicos.

Ordenador cuántico de uso específico. Ordenador cuántico (por ejemplo, un dispositivo cuantocriptográfico o una máquina cuántica de factorización) que ha sido diseñado para realizar una función determinada.

Discohercncia. Situación que se presenta cuando grupos de átomos que intervienen en un cálculo cuántico afectan de manera diferencial a otros átomos del entorno, lo que hace que la interferencia se reduzca y el cálculo pueda fallar. Es el principal obstáculo para la realización práctica de ordenadores cuánticos más potentes.

SUMARIO

Las leyes de la física permiten la existencia de ordenadores capaces de reproducir cualquier entorno físicamente posible sin utilizar para ello cantidades tan grandes de recursos que los hagan impracticables. Así pues, el cálculo universal no es solamente posible, como enuncia el principio de Turing, sino también tratable. Los fenómenos cuánticos pueden implicar vastos números de universos paralelos y, por consiguiente, ser difíciles de simular en un solo universo. No obstante, el concepto de universalidad se mantiene inalterado, porque los ordenadores cuánticos pueden reproducir de manera eficiente cualquier entorno cuántico físicamente posible, incluso cuando interactúan gran número de universos. Los ordenadores cuánticos pueden también resolver con eficiencia determinados problemas matemáticos, tales como la factorización, intratables de acuerdo con la física clásica, y pueden, asimismo, aplicar clases de criptografía según los métodos clásicos. El cálculo cuántico es un medio cualitativamente nuevo de

dominar la naturaleza.

El próximo capítulo puede resultar polémico para muchos matemáticos. Es inevitable. La matemática no es lo que ellos creen.

(Puede que los lectores no familiarizados con las asunciones tradicionales sobre la certidumbre del conocimiento matemático consideren que la conclusión principal del próximo capítulo —que nuestro conocimiento de la verdad matemática depende de nuestro conocimiento del mundo físico, y no es más fiable que éste— es obvia. Quizás esos lectores prefieran hojear someramente dicho capítulo y pasar al estudio del tiempo, que encontrarán en el capítulo 11.)

10. LA NATURALEZA DE LAS MATEMÁTICAS

La «estructura de la realidad» que he descrito hasta aquí ha sido la de la realidad *física.* No obstante, también me he referido libremente a entidades que no existen en el mundo físico, abstracciones tales como números y conjuntos infinitos de programas imformáticos. Las leyes de la física, por su parte, no son entidades físicas en el mismo sentido que las rocas y los planetas. Como he dicho, el «Libro de la Naturaleza» de Galileo es tan sólo una metáfora. Y no debemos olvidar las ficciones de la realidad virtual, los entornos inexistentes cuyas leyes difieren de las leyes de la física reales. Aún más allá se encuentran los que he denominado «entornos cantgotu», que no pueden ser reproducidos ni siquiera por la realidad virtual y que, como he dicho, existen en número infinito por cada entorno que puede ser reproducido. Pero ¿qué significa la afirmación de que dichos entornos «existen»? Si no se encuentran en la realidad, ni pueden ser reproducidos por la realidad virtual, ¿dónde existen?

¿*Existen* realmente las entidades abstractas, no físicas? Y, de ser así, ¿forman parte integrante de la estructura de la realidad? No me interesan las cuestiones meramente terminológicas: resulta obvio que los números, las leyes de la física y las demás entidades abstractas «existen» en algunos sentidos y no en otros. La verdadera pregunta *es:* ¿cómo debemos entender tales entidades? ¿Cuáles de ellas no son más que términos adoptados por mera conveniencia que, en última instancia, se refieren tan sólo a la realidad física ordinaria? ¿Cuáles son, simplemente, manifestaciones efímeras de nuestra cultura? ¿Cuáles son arbitrarias, semejantes a las reglas de un juego que debemos conocer para poder jugar a él? ¿Y cuáles, si es que las hay, sólo pueden ser explicadas de un modo que les atribuye existencia propia? Las de esta última categoría *deben* formar parte de la estructura de la realidad tal como la define este libro, ya que deberíamos entenderlas para poder comprender todo lo que es comprensible.

Esto parece sugerirnos que apliquemos de nuevo el criterio del doctor Johnson. Si queremos saber si una determinada abstracción existe realmente, debemos preguntarnos si es capaz de «devolver el golpe» de un modo complejo y autónomo. Los matemáticos, por ejemplo, caracterizan los «números naturales»

1, 2, 3, ..., en primera instancia, mediante una definición precisa, como ésta:

a) 1 es un número natural;

b) todo número natural tiene precisamente un siguiente, que es, a su vez, un número natural;

c) 1 no es el siguiente de ningún número natural;

d) si dos números naturales tienen el mismo siguiente, son el mismo número.

Tales definiciones son intentos de expresar de manera abstracta la noción intuitiva *física* de incrementos sucesivos de una cantidad discreta. (Más exactamente, y como expliqué en el capítulo anterior, esta noción es, en realidad, cuantomecánica.) Las operaciones aritméticas, como la suma y la multiplicación, al igual que otros conceptos, por ejemplo el de número primo, se definen con referencia a los «números naturales». Pero, a pesar de haber creado «números naturales» abstractos mediante esa definición, y de entenderlos gracias a esa intuición, nos encontramos con que nos queda mucho por comprender acerca de ellos. La definición de número primo fija definitivamente qué números son primos y cuáles no lo son, pero la *comprensión* de qué números son primos -por ejemplo, cómo se distribuyen los números primos en escalas muy grandes, cómo se agrupan, cuán «aleatorios» son y por qué- conlleva un mundo de nuevas reflexiones y explicaciones. Y, en efecto, la teoría de los números resulta ser un mundo en sí misma. Para comprender los números más plenamente, debemos definir muchas clases nuevas de entidades abstractas y postular numerosas estructuras nuevas, así como las conexiones entre ellas. Nos encontraremos con que algunas de esas estructuras abstractas están relacionadas con otras intuiciones anteriores que, aparentemente, no tenían ninguna relación con los números, como pueden ser la *simetría*, la *rotación*, el *continuo*, los *conjuntos*, el *infinito* y muchas más. Así pues, las entidades matemáticas abstractas que consideramos familiares pueden, en ocasiones, sorprendernos o contrariarnos. Pueden presentarse inesperadamente con nuevos aspectos, o disfrazadas. Pueden parecer inexplicables y más tarde, ajustarse a una nueva explicación. Son, pues, complejas y autónomas, y, por

consiguiente, debemos concluir, según el criterio del doctor Johnson, que son reales. Puesto que no podemos comprenderlas como formando parte de nosotros ni como formando parte de cualquier otra cosa que ya comprendamos, pero, en cambio, las *podemos* comprender como entidades independientes, debemos concluir que *son* entidades reales e independientes.

Sin embargo, las entidades abstractas son intangibles. No devuelven el golpe físicamente en el sentido en el que lo haría una piedra, de modo que la experimentación y la observación no pueden tener en las matemáticas el mismo papel que tienen en la ciencia. Dicho papel queda reservado en las matemáticas a la *demostración.* La piedra del doctor Johnson devolvió el golpe haciendo que su pie rebotara. Los números primos lo hacen cuando podemos demostrar algo inesperado sobre ellos, y, en especial, cuando, además, podemos explicarlo. Desde el punto de vista tradicional, la diferencia crucial entre demostración y experimento es que la primera no hace referencia alguna al mundo físico. Podemos realizar una demostración en la intimidad de nuestra mente o recluidos en el interior de un generador de realidad virtual que reproduzca una física falsa. Mientras sigamos las reglas de la inferencia matemática, deberíamos llegar a la misma respuesta que cualquier otra persona. Una vez más, la visión predominante es que -errores aparte-, cuando hemos demostrado algo que sabemos con *absoluta certeza*, *es* cierto.

Los matemáticos están orgullosos de esa certeza absoluta, y los científicos tienden a envidiarlos un poco por ello. Y es que en ciencia no hay manera de estar absolutamente seguro de ninguna proposición. Por muy bien que expliquen las teorías de que disponemos las observaciones actuales, en cualquier momento puede aparecer una nueva e inexplicable observación que cuestione la totalidad de nuestra estructura explicativa. Y, lo que es aún peor, alguien puede alcanzar una mejor comprensión que explique no tan sólo todas las observaciones actuales, sino también por qué las explicaciones anteriores parecían funcionar y, a pesar de ello, son erróneas. Galileo, por ejemplo, halló una nueva explicación para la antiquísima observación de que el suelo que pisamos está en reposo, una explicación que incluía el hecho de que el suelo, realmente, se mueve. La realidad virtual -que puede hacer que un entorno se parezca a otro— subraya el hecho de que, cuando la observación constituye el árbitro definitivo entre teorías, no puede existir jamás ni siquiera la remota seguridad de

que ninguna de las explicaciones de que disponemos, por obvia que parezca, sea cierta. Se supone, en cambio, que, cuando el árbitro es la demostración, existe la certeza.

Se dice que las reglas de la lógica fueron formuladas originalmente con la esperanza de que proporcionaran un método imparcial e infalible para la resolución de cualquier disputa. Esta esperanza nunca podrá verse justificada. El propio estudio de la lógica revela que el alcance de la deducción lógica como medio para el descubrimiento de la verdad presenta serias limitaciones. Dadas asunciones bien fundamentadas sobre el mundo, podremos deducir conclusiones, pero no podremos estar más seguros de éstas de lo que estábamos de aquéllas. Las únicas proposiciones que la lógica puede demostrar sin recurrir a asunciones son las tautológicas, asertos tales como «los planetas son planetas», que no afirman nada. En particular, todas las cuestiones fundamentales de la ciencia quedan fuera del ámbito en el que la lógica puede resolver disputas por sí sola. Se supone, sin embargo, que las matemáticas se hallan *dentro* de dicho ámbito. Los matemáticos buscan, pues, una verdad absoluta, pero abstracta, mientras que los científicos se consuelan con la idea de que pueden obtener un conocimiento fundamental y útil del mundo físico. Deben aceptar, sin embargo, que este conocimiento carece de garantías y será siempre provisional y falible. La idea de que la ciencia se caracteriza por la «inducción», método de justificación que se supone que es un equivalente algo falible de la deducción lógica, constituye un intento de sacar el mejor partido posible del status secundario que se atribuye al conocimiento científico. En lugar de aspirar a certezas justificadas por la deducción, quizás nos debamos conformar con cuasicertezas, certezas lo más aproximadas posible justificadas por la inducción.

Como ya he argumentado, no existe ningún método de justificación basado en la «inducción». La idea de que es posible razonar un camino hacia la «cuasicerteza» en ciencia es un mito. ¿Cómo podría asegurar con «cuasicerteza» que una nueva y maravillosa teoría física, que rebatirá todas mis más asunciones incuestionadas sobre la realidad, no será publicada mañana mismo? ¿O que no me encuentro en el interior de un generador de realidad virtual? Pero todo esto no equivale a decir que el conocimiento científico sea, efectivamente, de «segunda clase». Y es que la idea de que las matemáticas proporcionan certezas absolutas *es también un mito.*

Desde la antigüedad, la idea de que el conocimiento matemático tiene un status privilegiado ha ido asociada a menudo con la de que algunas entidades abstractas no sólo forman parte de la estructura de la realidad, sino que son incluso más reales que el mundo físico. Pitágoras creía que la regularidad que se observa en los procesos naturales era la expresión de las relaciones matemáticas entre los números naturales. «Todas las cosas son números», era su axioma, aunque no lo decía en sentido absolutamente literal. Platón, no obstante, fue más allá y negó la existencia real del mundo físico. Consideraba nuestras experiencias inconsistentes y engañosas, y argumentaba que los fenómenos y objetos físicos que percibimos son meras «sombras» o imitaciones imperfectas de sus esencias ideales («formas» o «ideas»), que existen en un dominio separado que constituye la verdadera realidad. En dicho dominio se encuentran, entre otras, las formas de los números puros, como 1, 2, 3, ..., y las de las operaciones matemáticas, como la suma y la multiplicación. Podemos percibir algunas sombras de estas formas; por ejemplo, cuando colocamos una manzana sobre una mesa, y luego otra, y decimos que hay dos manzanas. Pero las manzanas muestran su «un-idad» o su «dual-idad» (así como su «manzan-idad») tan sólo de manera imperfecta. No son idénticas en todo, de modo que nunca hay realmente *dos* sobre la mesa. Se podría objetar que el número dos puede también estar representado por la presencia de dos objetos *diferentes* sobre la mesa. Sin embargo, esta representación seguiría siendo imperfecta, porque habrá que admitir que algunas células pueden haberse desprendido de las manzanas y caer sobre la mesa, en la que también es fácil que estén presentes partículas de polvo, moléculas de aire, etcétera. A diferencia de Pitágoras, Platón no tenía ningún interés personal por los números naturales. Su realidad contenía las formas de todos los conceptos. Contenía, por ejemplo, la forma de un círculo perfecto. Los «círculos» que percibimos nunca son realmente círculos; no son del todo redondos ni planos, tienen siempre un espesor finito, etcétera. Todos ellos son imperfectos.

Platón había planteado así un problema: dada toda esta imperfección terrenal (y, hubiese podido añadir, dado nuestro imperfecto acceso sensorial incluso a los círculos terrenales), ¿cómo es posible que sepamos lo que sabemos acerca de los círculos reales y perfectos? Es evidente que sabemos que existen, pero ¿cómo? ¿De dónde sacó Eu-clides sus conocimientos de

geometría, expresados en sus famosos axiomas, si no tenía acceso a puntos, líneas rectas o círculos genui-nos? ¿De dónde viene la certeza de una demostración matemática, si nadie puede percibir las entidades abstractas a las que se refiere? La respuesta de Platón fue que no obtenemos nuestro conocimiento de esas entidades a partir de este mundo de sombra e ilusión, sino directamente del propio mundo de las formas. Según él, tenemos un conocimiento innato absolutamente perfecto del mundo real, pero lo olvidamos al nacer y lo vamos ocultando a lo largo de la vida superponiéndole capa tras capa de errores causados por la confianza en nuestros sentidos. Pero la realidad puede ser recordada mediante la diligente aplicación de la «razón», que al final acaba dándonos la certeza absoluta que la experiencia nunca podrá proporcionar.

Me pregunto si alguien (incluyendo al propio Platón, quien era, después de todo, un filósofo muy competente que creía que era bueno contarle al pueblo mentiras que lo ennoblecieran) se habrá creído jamás esta teoría tan mal hilvanada. No obstante, el problema que planteaba -el de cómo podemos tener no ya una certeza absoluta, sino un mero conocimiento, de las entidades abstractas- es bien real, y algunos elementos de la solución que propone han formado parte desde entonces de la teoría del conocimiento dominante. En particular, las ideas fundamentales de que los conocimientos matemático y científico provienen de *fuentes distintas*, y de que la fuente «especial» que alimenta las matemáticas les confiere una *certeza absoluta*, siguen siendo aceptadas sin la más mínima crítica por la práctica totalidad de los matemáticos. Hoy día se la denomina *intuición matemática*, pero dicha fuente sigue teniendo el mismo papel incuestionado que la «memoria» del mundo de las formas de Platón.

Ha habido controversias muy acerbas sobre qué clases precisas de conocimiento perfectamente fiable es de esperar que sean reveladas por la intuición matemática. En otras palabras, los matemáticos comparten la opinión de que la intuición matemática es una fuente de certeza absoluta, pero no pueden ponerse de acuerdo acerca de lo que les dice esa intuición. Obviamente, ello constituye un manantial inagotable de infinitas e irresolubles controversias.

Inevitablemente, buena parte de esas controversias se ha centrado en la validez o invalidez de diversos métodos de demostración. La causa de una de ellas fueron los números

llamados «imaginarios», que son raíces cuadradas de números negativos. Nuevos teoremas acerca de los números ordinarios o «reales» eran demostrados -en estadios intermedios de la demostración— mediante referencias a las propiedades de los números imaginarios. Los primeros teoremas sobre la distribución de los números primos, por ejemplo, fueron demostrados de ese modo. Algunos matemáticos plantearon objeciones a los nú-leeros imaginarios aduciendo que no eran reales. (La terminología corriente refleja aún esa antigua controversia, si bien en la actualidad se considera que los números imaginarios son tan reales como los «reales».) Supongo que los profesores de esos matemáticos les dijeron que no les estaba *permitido* sacar la raíz cuadrada de menos uno, de modo que pensaron que a nadie más le estaba permitido hacerlo. Sin duda, llamaron a este impulso poco caritativo «intuición matemática». Pero otros matemáticos tenían otras intuiciones. Comprendieron qué eran los números imaginarios y cómo encajaban con los números reales. ¿Por qué, pensaron, no podemos definir nuevas entidades abstractas y otorgarles las propiedades que creamos convenientes? Sin duda, la única base legítima que lo impedía era que esas propiedades fuesen lógicamente inconsistentes. (Este es, en esencia, el consenso moderno, que el matemático John Horton Conway denomina, con cierta exageración, «Movimiento de Liberación de los Matemáticos».) Es cierto que nadie ha demostrado que el sistema de números imaginarios *sea* autoconsistente, pero, si bien se mira, tampoco ha *demostrado* nadie que la aritmética común de los números naturales lo sea.

Hubo controversias similares acerca de la validez de la utilización de números infinitos, de las series con un número infinito de elementos y de las cantidades infinitesimales utilizadas en el cálculo. David Hilbert, el gran matemático alemán que proporcionó buena parte de la infraestructura matemática de la teoría general de la relatividad y la teoría cuántica, subrayó que la «bibliografía matemática, está abarrotada de sandeces y disparates que tuvieron su origen en el infinito». Y es que algunos matemáticos, como veremos, negaban toda validez al razonamiento sobre entidades infinitas. El avance espectacular de la matemática pura en el transcurso del siglo XIX contribuyó bien poco a resolver estas controversias. Al contrario, tendió más bien a intensificarlas y a hacer que surgieran nuevas polémicas. A medida que el razonamiento matemático se volvía cada vez más

complejo, era inevitable que se alejara de la intuición, lo que tuvo dos importantes y contradictorios efectos. En primer lugar, los matemáticos se volvieron más exigentes para con las demostraciones, que se vieron sujetas a normas cada vez más estrictas para ser aceptadas. Y, por otro lado, se idearon nuevos y más perfectos métodos de comprobación, que, sin embargo, no siempre podían ser validados por los procedi-fc imiemos existentes. Todo ello hizo que, a menudo, se plantearan dudas acerca de si un determinado método de comprobación, por auto-. evidente que pareciese, era completamente infalible o no.

Por todo ello, hacia 1900 hubo una crisis en los fundamentos de las matemáticas causada, precisamente, porque se percibía que carecían de fundamentos. Pero ¿qué había sido de las leyes de la lógica pura? ¿No se suponía que debían ser capaces de resolver toda disputa dentro del ámbito de las matemáticas? La embarazosa realidad era que la causa de las disputas matemáticas eran, justamente, esas «leyes de la lógica pura». Aristóteles fue el primero en codificarlas, en el siglo IV a.C., y fundó así lo que se denomina hoy la *teoría de la demostración.* Llegó a la conclusión de que la demostración debía consistir en una secuencia de enunciados que empezara con varias premisas y definiciones y terminara con la conclusión deseada. Para que una secuencia de enunciados constituyese una demostración válida, cada uno de ellos, además de ajustarse a las premisas que lo originaban, debía hacerlo también a una de las diversas fórmulas de demostración denominadas *silogismos.* Un ejemplo de silogismo sería:

En otras palabras, dicha regla decía que, cuando un enunciado de la clase «todo A tiene la propiedad B» (como en «Todos los hombres son mortales») y otro de la clase «el individuo X es A» (como en «Sócrates es hombre») aparecen en una demostración, el enunciado «X tiene la propiedad B» («Sócrates es mortal») puede aparecer después válidamente en ella y, lo que es más, constituir una conclusión válida. Los silogismos expresaban lo que podríamos denominar *reglas de inferencia*, es decir, reglas que definen los pasos autorizados en las demostraciones para conseguir que la verdad de las premisas se transmita a las conclusiones. De acuerdo con el mismo principio, son reglas que pueden ser aplicadas para determinar si una demostración

propuesta es válida o no.

Aristóteles afirmó que toda demostración válida podía ser expresada en forma de silogismo, pero no lo demostró. A ello se añadía que la teoría de la demostración se enfrentaba al problema de que muy pocas demostraciones matemáticas modernas estaban expresadas puramente como secuencias de silogismos, y todavía eran menos las susceptibles de ser reformuladas en dicha forma, ni siquiera en principio. Y, además, la mayoría de los matemáticos no se resignaban a seguir la ley aristotélica al pie de la letra, puesto que muchas de las nuevas demostraciones parecían tan autoevidentemente válidas como el propio razonamiento de Aristóteles. Era evidente que las matemáticas habían avanzado. Nuevos instrumentos, como la lógica simbólica y la teoría de conjuntos, permitían a los matemáticos establecer nuevas relaciones entre las estructuras matemáticas. Ello había ido creando verdades autoevidentes que ya no dependían de las reglas clásicas de inferencia, de modo que éstas eran cada vez más autoevidentemente inadecuadas. Pero ¿cuáles de los nuevos métodos de demostración eran infalibles de verdad? ¿Cómo debían ser modificadas las reglas de inferencia para que pudiesen tener de una manera real la completitud que Aristóteles, equivocadamente, proclamaba? ¿Cómo podía ser recuperada la absoluta autoridad de las antiguas reglas cuando los matemáticos no lograban ponerse de acuerdo sobre qué era autoevidente y qué no era más que simples disparates?

Mientras tanto, los matemáticos continuaban levantando abstractos castillos en el aire. A efectos prácticos, muchas de esas construcciones parecían lo bastante sólidas. Algunas habían llegado a ser indispensables para la ciencia y la tecnología, y la mayoría estaban conectadas por una estructura explicativa satisfactoria y fructífera. Sin embargo, nadie podía garantizar que esa estructura, o buena parte de ella, no estuviese fundada sobre una contradicción lógica que pudiera acabar convirtiéndola, literalmente, en un disparate. En 1902 Bertrand Russell demostró que un esquema para definir rigurosamente la teoría de conjuntos, que acababa de ser presentado por el lógico alemán Gottlob Frege, era inconsistente. Ello no significaba que fuese necesariamente incorrecto utilizar los conjuntos en las demostraciones. De hecho, muy pocos matemáticos consideraron seriamente la posibilidad de que cualquiera de los modos habituales de utilizar los conjuntos, la aritmética o las otras ramas básicas de las matemáticas pudiera

no ser válido. El resultado al que llegó Russell causó desazón porque los matemáticos consideraban que su disciplina era el medio por excelencia para llegar a certezas absolutas mediante las demostraciones proporcionadas por los teoremas matemáticos. Que hubiese una posibilidad de controversia sobre la validez de diferentes métodos de demostración minaba todo el propósito (tal como era concebido) de la disciplina matemática.

Muchos matemáticos, pues, sentían que era urgente dotar a la teoría de la demostración —y, por lo tanto, a las propias matemáticas— de una base segura. Querían consolidar los progresos realizados y definir de una vez por todas qué clases de demostración eran absolutamente seguras y cuáles no. Lo que quedara fuera de la zona de seguridad podría ser desdeñado, y lo que permaneciera dentro de ella constituiría la única base válida para las matemáticas futuras.

Con este fin, el matemático holandés Luitzen Egbertus Jan Brouwer defendió una estrategia extremadamente conservadora para la teoría de la demostración, conocida como *intuicionismo*, que aún tiene algunos adeptos. Los intuicionistas tratan de interpretar la «intuición» del modo más restrictivo concebible, y sólo conservan de ella lo que consideran sus aspectos autoevidentes más incuestionables. Y elevan la intuición matemática así definida a un status más elevado que el que le otorgó Platón, ya que la sitúan en un plano superior incluso al de la lógica pura, ya que ésta les parece indigna de confianza, excepto cuando se halle justificada por la intuición matemática directa. Los intuicionistas niegan, por ejemplo, que sea posible la intuición directa de cualquier entidad infinita. Niegan, por consiguiente, la existencia de cualquier conjunto infinito, como el de los números naturales. Consideran autoevidentemente falsa la proposición «existen infinitos números naturales». Por la misma razón, la proposición «existen más entornos cantgotu que entornos físicamente posibles» les parecería carente del menor sentido.

Históricamente, el intuicionismo —al igual que el inductivismo— tuvo un valioso papel liberador. Se atrevió a cuestionar las certidumbres recibidas, algunas de las cuales resultaban, en efecto, falsas. Pero resulta inútil como teoría positiva sobre lo que es o no es una demostración matemática válida. En realidad, el intuicionismo es precisamente el equivalente matemático del solipsismo. En ambos casos se trata de una

reacción exagerada ante la idea de que no podemos estar seguros de conocer el amplio mundo que nos rodea. Ambas doctrinas proponen como solución la retirada a un mundo interior que se supone que podemos conocer de manera directa y, por consiguiente, fiel. Y, también en las dos, esa renuncia imposibilita la explicación de gran parte de lo que queda en el interior de ese ámbito privilegiado. Por ejemplo, si es falso —como defienden los intuicionistas— que haya infinitos números naturales, podemos inferir que tan sólo existe una cantidad finita de dichos números. ¿Cuántos? Y, por otra parte, por muchos que haya, ¿por qué no podemos intuir un número natural siguiente al último intuido? Los intuicionistas se quitarían el problema de encima señalando que la argumentación que acabo de ofrecer da por sentada la validez de la lógica ordinaria y, en particular, que parte de inferir, a partir del hecho de que no hay infinitos números naturales, que un número debe ser finito. La principal regla de inferencia es la denominada *ley del tercero excluido*, la cual dice que, dada cualquier proposición X (como «hay infinitos números naturales»), o bien X es cierta o bien lo es su negación («hay una cantidad finita de números naturales»); no existe una tercera posibilidad. Los intuicionistas niegan impertérritos la ley del tercero excluido.

Puesto que en la mente de la mayoría de las personas la ley del tercero excluido está respaldada por una fuerte intuición, su negación provoca que los no intuicionistas se pregunten si, después de todo, la intuición de los intuicionistas será tan autoevidentemente fiable como proclaman. Por otro lado, si consideramos que la ley del tercero excluido surge de una intuición *lógica*, ello nos conducirá a reexaminar la cuestión de si la intuición matemática sustituye realmente a la lógica. En cualquier caso, ¿será *autoevidente* que lo haga?

Por supuesto, todo lo dicho hasta aquí no es más que una crítica del intuicionismo desde el exterior. No es una refutación, puesto que el intuicionismo, como el solipsismo, jamás podrá ser refutado. Si alguien insiste en que una determinada proposición autoconsistente es autoevidente para él, al igual que si asegura que sólo existe él, no podremos demostrar que está equivocado. Sin embargo, y como ocurre, en general, con el solipsismo, el fallo fundamental del intuicionismo no se pone de manifiesto cuando se le ataca, sino cuando se le toma en serio, de acuerdo con sus propios términos, como explicación de su propio mundo arbitrariamente truncado. Los intuicionistas creen en la realidad de

los números finitos naturales, 1, 2, 3, ..., e incluso 10.949.769.651.859. Sin embargo, el argumento intuitivo de que, puesto que cada uno de estos números tiene un siguiente, forman una secuencia infinita, no es, desde la perspectiva intuicionista, más que un autoengaño o ficción, por lo que resulta, literalmente, insostenible. Pero al cortar el vínculo entre su versión de los «números naturales» abstractos y las intuiciones para cuya formalización estaban inicialmente destinados estos números, los intuicionistas se niegan a sí mismos la habitual estructura explicativa mediante la cual pueden ser entendidos los números naturales. Esto presenta un problema para cualquiera que prefiera las explicaciones a las complicaciones inexplicadas. En lugar de resolver el problema proporcionando una estructura explicativa alternativa o más profunda para los números naturales, el intuicionismo hace exactamente lo mismo que hacía la Inquisición y que hacen los solipsistas: se aparta aún más de la explicación e introduce complicaciones inexplicadas adicionales (en este caso, la negación de la ley del tercero excluido), cuyo único propósito es permitir que los intuicionistas se comporten como si la explicación de sus oponentes fuese cierta, pero sin sacar de ello ninguna conclusión sobre la realidad.

Al igual que el solipsismo se inicia con la motivación de simplificar un mundo incierto y sobrecogedoramente diverso, pero cuando se le toma en serio resulta no ser otra cosa que realismo *más* algunas complicaciones innecesarias, el intuicionismo acaba por ser una de las doctrinas más contraintuitivas que jamás hayan sido defendidas seriamente.

David Hilbert propuso un plan mucho más sensato —aunque también terminó en fracaso— para «establecer de una vez por todas la certeza de los métodos matemáticos». El plan de Hilbert se basaba en el concepto de consistencia. Aspiraba a fijar de manera definitiva una serie completa de reglas modernas de inferencia para las demostraciones matemáticas, dotadas de ciertas propiedades. Serían finitas en número. Serían directamente aplicables, de modo que determinar si una demostración propuesta las satisfacía o no constituyese un ejercicio incontrovertible. De preferencia, deberían ser intuitivamente autoevi-dentes, si bien ésta no constituía una condición excluyente para el pragmático Hilbert. Se habría dado por satisfecho si las reglas se hubiesen correspondido tan sólo moderadamente bien con la intuición, a condición de poder estar seguro de que fuesen autoconsistentes.

En otras palabras, quería estar seguro de que si las reglas designaban como válida a una determinada demostración, no pudiesen jamás hacer lo mismo con otra demostración que arrojara una conclusión contraria. ¿Cómo esperaba conseguirlo? Esta vez, la consistencia tenía que quedar *demostrada* mediante un método coherente, a su vez con las propias reglas de inferencia. Hilbert esperaba que, de este modo, la completitud y la certeza aristotélicas serían restauradas y todo enunciado matemático verdadero podría, en principio, ser demostrable con dichas reglas, mientras que ningún enunciado falso podría serlo. En 1900, con motivo del II Congreso Internacional de Matemáticas celebrado en París, Hilbert presentó una relación de los veintitrés problemas que esperaba que los matemáticos fuesen capaces de resolver durante el transcurso del siglo XX. En décimo lugar figuraba el problema de encontrar la serie de reglas de inferencia con las propiedades mencionadas, y de demostrar su consistencia de acuerdo con sus propios estándares.

Pero Hilbert vio contrariadas sus esperanzas. Treinta y un años después, Kurt Gödel revolucionó la teoría de la demostración con una completa refutación de la que los mundos matemático y filosófico aún no se han recuperado: demostró que el décimo problema de Hilbert es insoluble. Para empezar, Gödel demostró que ningún conjunto de reglas de inferencia, aunque fuese capaz de validar correctamente las demostraciones de la aritmética común, podría validar jamás la demostración de su propia consistencia. No había, por consiguiente, esperanza alguna de encontrar la serie demostrablemente consistente de reglas que propuso Hilbert. En segundo lugar, demostró que si un conjunto de reglas de inferencia de alguna rama (lo suficientemente bien dotada) de las matemáticas *es* consistente (demostrablemente o no), dentro de dicha rama deben existir métodos válidos de demostración cuya validez no pueden demostrar dichas reglas. Es lo que se conoce como *teorema de Godel de la incompletitud.* Para demostrar sus teoremas, Godel utilizó una notable extensión del «argumento diagonal» de Cantor, que he mencionado en el capítulo 6. Empezó por considerar cualquier conjunto consistente de reglas de inferencia, para demostrar después cómo formular una proposición que no pudiese ser ni demostrada ni refutada según dichas reglas. Finalmente, demostró que dicha proposición sería cierta.

Si el programa de Hilbert hubiese tenido éxito, ello habría

sido una mala noticia para el concepto de realidad que propongo en este libro, ya que habría eliminado la necesidad de la *comprensión* al juzgar ideas matemáticas. Cualquiera —o cualquier máquina debidamente programada— que se aprendiese de memoria las reglas de inferencia que Hilbert tanto ansiaba hallar, estaría en las mismas condiciones que el más capacitado matemático para ser juez de cualquier proposición matemática, si bien no necesitaría la percepción o comprensión de aquél, ni tampoco tener la más remota idea de los problemas planteados por dichas proposiciones. En principio, sería posible hacer nuevos descubrimientos matemáticos sin saber más matemáticas que las reglas de Hilbert. Sólo habría que comprobar todas las series de letras y símbolos matemáticos por orden alfabético, hasta que alguno de ellos superara la prueba de ser la demostración o la refutación de alguna conjetura famosa sin resolver. En principio, se podría dilucidar cualquier controversia matemática incluso sin entenderla ni conocer siquiera el significado de los símbolos, por no hablar del funcionamiento de la demostración, o de lo que quería demostrar, o de qué método de demostración se trataba, o de por qué era fiable.

A primera vista, parece que el logro de un estándar unificado de demostración matemática habría sido, al menos, una ayuda para la tendencia general a la unificación, es decir, para la «profundización» de nuestro conocimiento a la que me refería en el capítulo 1; pero, en realidad, habría ocurrido todo lo contrario. Como en el caso de la predictiva «teoría total» en física, las reglas de Hilbert no nos habrían dicho casi nada sobre la estructura de la realidad. Como mucho, y por lo que concierne a las matemáticas, habrían hecho realidad la visión última reduccionista: predecir todo (en principio), pero no explicar nada. Es más, si las matemáticas hubiesen resultado ser reduccionistas, estarían presentes en ellas todas las características indeseables que, según argumenté en el capítulo 1, no forman parte de la estructura del conocimiento humano: las ideas matemáticas habrían constituido una jerarquía, con las reglas de Hilbert como base. Las verdades matemáticas cuya verificación de acuerdo con dichas reglas fuese demasiado compleja se habrían visto relegadas a la condición de objetivamente menos fundamentales que aquellas comprobables de modo inmediato según esas reglas. Puesto que sólo habría habido un repertorio limitado de semejantes verdades fundamentales, con el paso del tiempo las matemáticas habrían

tenido que tratar cada vez menos problemas fundamentales. Las matemáticas habrían llegado, según esta sombría hipótesis, a un punto final. En caso contrario, se habrían ido fragmentando inevitablemente en especialidades cada vez más misteriosas a medida que fuese aumentando la complejidad de los temas «emergentes» que los matemáticos se viesen obligados a analizar y que las conexiones entre dichos temas y los fundamentos de la disciplina resultasen cada vez más remotas.

Gracias a Gödel sabemos que no existirá jamás un método fijo para determinar si una proposición matemática es irrefutable, del mismo modo que no lo hay para decidir si es cierta una teoría científica. Tampoco habrá nunca un método fijo para crear nuevos conocimientos matemáticos. Por consiguiente, el progreso de las matemáticas dependerá siempre del ejercicio de la creatividad. En todo momento les resultará posible —y necesario— a los matemáticos inventar nuevas clases de demostración. Las validarán mediante nuevas argumentaciones y nuevas formas de explicación, que dependerán de la mejora constante de su comprensión de las entidades abstractas involucradas. Los propios teoremas de Gödel son un buen ejemplo de ello: para demostrarlos, tuvo que inventar un nuevo método. Ya dije que dicho método se basaba en el «argumento diagonal», pero Gödel lo extendió y le dio una nueva forma. Nada había sido demostrado mediante dicha forma de argumentación con anterioridad, *y* ninguna regla de inferencia establecida previamente por alguien que no conociese el método de Gödel hubiese podido predecir nunca que era válido. Y, sin embargo, *es* autoevidentemente válido. ¿De dónde proviene esa autoevidencia? Pues de que Gödel comprendió la naturaleza de la demostración. Las demostraciones de Gödel son tan convincentes como cualesquiera otras en matemáticas, pero sólo si se comprende previamente la explicación que las acompaña.

Así pues, la explicación desempeña, después de todo, el mismo papel crucial en las matemáticas que en la ciencia. El objeto final es en ambos casos la explicación y la comprensión del mundo, tanto del mundo físico como del mundo de las abstracciones matemáticas. La demostración y la observación son, simplemente, medios con los que comprobamos nuestras explicaciones.

Roger Penrose ha sacado una conclusión adicional, radical y muy platónica, de los resultados de Gödel. Como Platón, Penrose

se siente fascinado por la capacidad de la mente humana para captar las certezas abstractas de las matemáticas. Pero, a diferencia de aquél, Penrose no cree en lo sobrenatural y da por sentado que el cerebro es parte del mundo natural y sólo tiene acceso a él. Por ello, el problema resulta más complejo para él de lo que lo era para Platón: ¿cómo puede el borroso e incierto mundo físico ofrecer certezas matemáticas a esa borrosa e incierta parte de sí mismo que es un matemático? En particular, Penrose se maravilla de que podamos percibir realmente la infalibilidad de nuevas y válidas *formas* de demostración, de las que, según asegura Gödel, existen reservas ilimitadas.

Penrose aún busca una respuesta detallada, pero reconoce que la existencia de esa especie de intuición matemática sin límites es fundamentalmente incompatible con la estructura actual de la física y, en particular, con el principio de Turing. Su argumentación se desarrolla, en resumen, como sigue: si el principio de Turing es cierto, podemos considerar que el cerebro (como cualquier otro objeto) es un ordenador que ejecuta un determinado programa. Las interacciones del Cerebro con el entorno constituyen los datos y resultados del progra-
Consideremos ahora a un matemático en el acto de decidir si una determinada clase de demostración recientemente propuesta es o no válida. Tomar esta decisión equivale a ejecutar un programa de validación de demostraciones en el interior de la mente del matemático. Dicho programa incorpora una serie de reglas de inferencia de

Hilbert que, según el teorema de Gödel, no puede ser completa. Es más, y como ya he dicho, Gödel nos proporciona el modo para formular y demostrar una proposición verdadera, que dichas reglas nunca podrán reconocer como demostrada. Por consiguiente, nuestro matemático, cuya mente es, en realidad, un ordenador que aplica esas reglas, tampoco podrá nunca reconocer la proposición como demostrada. Penrose propone demostrar entonces la proposición, así como el método de Gödel que demuestra que es cierta, a ese matemático. Éste entenderá la demostración, que, después de todo, es válida de un modo autoevidente, por lo que el matemático puede, presumiblemente, comprender que lo es. Pero ello contradeciría el teorema de Gödel, de modo que debe haber una falsa asunción en algún lugar de la argumentación. Penrose cree que el principio de Turing es la falsa asunción.

La mayoría de los científicos informáticos no están de acuerdo con Penrose en que el principio de Turing sea el eslabón más débil de su razonamiento. Podrían argumentar que es posible que nuestro matemático sea, efectivamente, incapaz de reconocer la proposición gódeliana como demostrada. Quizás parezca extraño que un matemático se vuelva de repente incapaz de comprender una demostración autoevidente, pero echemos un vistazo a esta proposición: *David Deutsch no puede juzgar consistentemente que este enunciado sea cierto.* Por más que lo intente, no puedo juzgar consistentemente que sea cierto. Si lo hiciera, estaría juzgando que *no puedo* juzgar que sea cierto, con lo que me estaría contradiciendo. Pero *usted* puede ver que es cierto ¿no es así? Ello demuestra que, al menos, es posible que una proposición sea insoluble para una persona y, al mismo tiempo, autoevidentemente cierta para todas las demás.

En cualquier caso, Penrose confía en que una nueva y fundamental teoría de la física sustituya a la vez a la teoría cuántica y a la teoría general de la relatividad. Proporcionaría predicciones nuevas y comprobables, si bien estaría, por supuesto, de acuerdo con ambas teorías en lo concerniente a todas las observaciones existentes. (No se conocen demostraciones experimentales que contradigan dichas teorías.) Sin embargo, el mundo de Penrose es, en lo fundamental, muy distinto del que describe la física actual. Su estructura básica de la realidad es lo que *nosotros* denominamos el mundo de las abstracciones matemáticas. En este aspecto, Penrose, cuya realidad incluye todas las abstracciones matemáticas -pero quizás no *todas* las abstracciones (como el honor y la justicia)-, se encuentra en algún lugar intermedio entre Platón y Pitágoras. Lo que nosotros denominamos el mundo físico es para él completamente real (otra diferencia con Platón), pero, de algún modo, forma parte, o surge, de las propias matemáticas. Además, no existe la universalidad, y, en particular, no hay máquina que pueda representar todos los procesos del pensamiento humano. Sin embargo, el mundo (y, en especial, por supuesto, su sustrato matemático) sigue siendo comprensible. Su comprensibilidad ya no está asegurada por la universalidad del cálculo, sino por un fenómeno que resulta nuevo para la física (pero no para Platón): *las entidades matemáticas inciden directamente sobre el cerebro humano* mediante procesos físicos aún por descubrir. De este modo, el cerebro, según Penrose, no hace matemáticas tan sólo por referencia a lo que

corrientemente denominamos el mundo físico, sino que tiene acceso directo a la realidad platónica de las formas matemáticas, en la que puede percibir verdades matemáticas con (meteduras de pata aparte) absoluta certeza.

Se dice a menudo que el cerebro tal vez sea un ordenador cuántico, cuyas intuiciones, conciencia y capacidad para resolver problemas podrían basarse en cálculos cuánticos. Esto *podría* ser así, pero no conozco ninguna evidencia ni ninguna argumentación convincente de que lo sea. Mi opinión es que si el cerebro es un ordenador, es un ordenador clásico. Pero esta cuestión es independiente de las ideas de Penrose, quien no defiende que el cerebro sea una nueva clase de ordenador universal que sólo se diferenciaría del ordenador universal cuántico por su mayor repertorio de cálculos, posible gracias a la nueva física poscuántica. Lo que argumenta es una nueva física, que ya no apoyará la universalidad del cálculo, de modo que, según esa nueva teoría, ya no será posible considerar ciertas acciones del cerebro como cálculos.

Debo admitir que me resulta imposible concebir una teoría así. No obstante, los avances fundamentales suelen ser difíciles de concebir antes de que ocurran. Por descontado, no es fácil juzgar la teoría de Penrose antes de que consiga formularla por completo. En la eventualidad de que una teoría con las propiedades que él busca sustituyera a la teoría cuántica, o a la general de la relatividad, o a ambas —bien mediante la comprobación experimental, bien proporcionando Un nivel más profundo de explicación—, toda persona razonable desearía adoptarla. Nos embarcaríamos entonces en la aventura de comprender la nueva concepción del mundo que las estructuras explicativas de la teoría nos inducirían a adoptar. Parece probable que fuera una concepción del mundo muy distinta de la que presento en este libro. No obstante, aunque todo esto llegara a ocurrir, sigo sin ser capaz de ver cómo la motivación original de la teoría -la de explicar nuestra capacidad de captar nuevas demostraciones matemáticas-, podría quedar satisfecha. Seguiría siendo un hecho que, tanto hoy como a lo largo de la historia, los grandes matemáticos han tenido distintas y conflictivas percepciones de la validez de diversos métodos de demostración. Así pues, aun en el caso de que fuese cierto que una realidad psicomatemática absoluta pudiese alimentar directamente nuestros cerebros para crear intuiciones matemáticas, estaría claro que los matemáticos

no siempre serían capaces de distinguir esas intuiciones de otras ideas e intuiciones erróneas. Por desgracia, no suena un timbre o se enciende una luz cada vez que comprendemos una demostración auténticamente válida. Es muy probable que en más de un momento gritásemos *«¡Eureka!»* y, sin embargo, estuviéramos equivocados. Incluso si la teoría predijera que *existe* algún indicador físico, no detectado anteriormente, que acompaña a las verdaderas intuiciones (lo cual, hoy por hoy, parece muy inverosímil), aunque podría resultarnos útilísimo, seguiría sin ser el equivalente de una demostración de que el indicador funcionaría. Nada podría demostrar que una teoría todavía mejor que la de Penrose no pudiera algún día sustituirla, poner de manifiesto que el supuesto indicador no era fiable, después de todo, y proponer otro mejor. De modo que, aun haciendo todas las posibles concesiones a la proposición de Penrose, áun imaginando que sea cierta y contemplando el mundo enteramente en sus términos, seguimos sin poder explicar la pretendida certeza del conocimiento adquirido por la práctica de las matemáticas.

He presentado tan sólo un esquema de las argumentaciones de Penrose y sus oponentes. El lector habrá notado que, esencialmente, me sumo a estos últimos. Sin embargo, aun dando por sentado que el argumento gódeliano de Penrose no consigue demostrar lo que desea, y que la nueva teoría física que propone no parece capaz de ser una explicación válida, no por ello deja Penrose de tener razón cuando dice que toda concepción del mundo basada en la idea actual de la racionalidad científica crea un problema para los fundamentos generalmente aceptados de las matemáticas (o, como lo expondría él, a la inversa). Se trata de un viejo problema, que ya planteó Platón, un problema que, como señala Penrose, se acentúa a la luz tanto del teorema de Gödel como del principio de Turing, y que es el siguiente: en una realidad basada en la física y comprendida mediante los métodos científicos experimentales, ¿de dónde procede la certeza matemática? Aunque la mayoría de los matemáticos y los científicos informáticos dan por sentada la certidumbre de la intuición matemática, no se toman en serio el problema de conciliaria con la visión científica. Penrose sí lo hace, y propone una solución. Ésta contempla un mundo comprensible, rechaza lo sobrenatural, reconoce que la creatividad es fundamental para las matemáticas, confiere realidad objetiva tanto al mundo físico como a las entidades abstractas e implica una integración de las bases

de las matemáticas y las de la física. En todos estos aspectos, coincido con él.

Puesto que los intentos de Brouwer, Hilbert, Penrose y otros matemáticos para superar el reto de Platón no parecen haber tenido éxito, merece la pena volver a analizar la aparente refutación realizada por Platón de la idea de que la verdad matemática puede ser obtenida mediante métodos científicos.

En primer lugar, Platón dice que, puesto que sólo tenemos acceso a círculos imperfectos, no podemos, en consecuencia, obtener ningún conocimiento sobre los círculos perfectos. Pero ¿por qué no, exactamente? De acuerdo con esa argumentación, podría decirse que no podemos descubrir las leyes del movimiento de los planetas porque no tenemos acceso directo a ellos, sino tan sólo a sus imágenes (esto es lo que *decía* la Inquisición, y he demostrado ya en qué se equivocaba). Podría argumentarse también que es imposible fabricar máquinas herramientas precisas, porque la primera debería por fuerza ser fabricada con máquinas herramientas imprecisas. La perspectiva que tenemos en la actualidad nos permite ver que esa línea de crítica se basa en una imagen muy burda del modo en que funciona la ciencia —semejante, hasta cierto punto, al inductivismo—, lo cual no debería sorprendernos, ya que Platón vivió con anterioridad a la aparición de lo que ahora llamamos ciencia. Si, por ejemplo, la única manera de aprender algo sobre los círculos mediante la experiencia fuera examinar miles de círculos reales y tratar después, a partir de los datos acumulados, de inferir algo sobre sus homólogos abstractos euclídeos, Platón estaría en lo cierto. Pero si formulamos la hipótesis de que los círculos reales se parecen a los abstractos en determinados aspectos, y resulta que acertamos, podremos aprender algo sobre los círculos abstractos mediante la observación de los reales. En la geo-Hietría euclídea se emplean a menudo diagramas para enunciar un determinado problema o su solución. Este método descriptivo podría toducir a errores en el caso de que los círculos del diagrama tuviesen imperfecciones que Ies dieran una apariencia engañosa, por ejemplo si dos círculos pareciesen tocarse aunque en realidad no tuviesen ningún punto de contacto. Pero si se entiende la relación entre los círculos reales y los perfectos, se puede, con cuidado, eliminar semejantes errores. En cambio, si no se entiende esa relación, resultará prácticamente imposible comprender la geometría euclídea.

La fiabilidad del conocimiento de un círculo *perfecto* que podemos conseguir mediante el *diagrama* de un círculo depende por completo de la exactitud de la hipótesis de que ambos se parecen en los aspectos relevantes. Semejante hipótesis, pues, referida a un objeto físico (el diagrama), equivale a una teoría física, y, como tal, nunca podrá ser conocida con absoluta certeza. Pero ello no excluye, como diría Platón, la posibilidad de aprender acerca de los círculos perfectos mediante la experiencia. Lo único que excluye es la posibilidad de la certeza, lo cual no debería preocupar a quien busque explicaciones y no certidumbre.

La geometría euclídea puede ser formulada en su totalidad sin recurrir a diagramas, pero el modo en que números, letras y símbolos matemáticos son utilizados en la demostración simbólica no puede generar mayor certidumbre que un diagrama, y por la misma razón. Los símbolos son también objetos físicos -rasgos escritos con tinta sobre papel, por ejemplo- que denotan objetos abstractos. Una vez más, nos apoyamos enteramente en la hipótesis de que el comportamiento físico de los símbolos se corresponde con el de las abstracciones que representan. Por consiguiente, la fiabilidad de lo que aprendamos mediante la manipulación de dichos símbolos se basará por completo en la precisión de nuestras teorías sobre su comportamiento físico, así como en el comportamiento de nuestras manos, ojos y demás órganos con los que los manipulamos y observamos. Una tinta trucada, que modificase el aspecto del símbolo cuando no lo estuviésemos mirando —quizás accionada por control remoto por algún bromista experto en alta tecnología—, podría hacer que nos engañásemos acerca de lo que tenemos por «seguro».

Reexaminemos ahora otra de las asunciones de Platón: la de que no podemos encontrar la perfección en el mundo físico. Tal vez tuviese razón en que no encontraremos en él el honor y la justicia perfectos, y, sin duda, la tenía en que no podremos encontrar en él las leyes de la física o la serie de todos los números naturales. Pero sí podemos encontrar una mano perfecta en bridge, o el movimiento perfecto en una determinada posición de ajedrez. Es decir, podemos hallar objetos o procesos físicos que posean plenamente las propiedades de las abstracciones especificadas. Podemos aprender ajedrez igual de bien sobre un tablero real que con la forma perfecta de un juego de ajedrez. El hecho de que un alfil esté desportillado no hace que el jaque mate sea menos terminante.

En realidad, es *posible hacer* que un círculo euclídeo perfecto sea accesible a nuestros sentidos. Platón no lo podía saber, porque aún no se conocía la realidad virtual. No resultaría particularmente difícil programar los generadores de realidad virtual que planteé en el capítulo 5 con las reglas de la geometría euclídea, de tal modo que el usuario pudiera experimentar una interacción con un círculo perfecto. Al carecer de grosor, el círculo resultaría invisible, a menos que modificáramos también las leyes de la óptica, y en este caso lo podríamos dotar de un resplandor que permitiera al usuario saber dónde se encuentra. (Los puristas quizás preferirían prescindir de estos adornos.) Podríamos hacer el círculo rígido e impenetrable, para que el usuario pudiese comprobar sus propiedades mediante la utilización de herramientas e instrumentos de medición igualmente rígidos e impenetrables. Los calibradores de los generadores de realidad virtual deberían alcanzar una precisión absoluta, para poder medir con exactitud el grosor cero. El usuario podría «dibujar» otros círculos u otras figuras geométricas, de acuerdo con las reglas de la geometría euclídea. Las dimensiones de las herramientas, así como las del propio usuario, podrían ajustarse a voluntad para permitir la comprobación a cualquier escala, por pequeña que fuera, de las predicciones de los teoremas geométricos. En todos los aspectos el círculo representado se ajustaría con absoluta precisión a lo especificado por los axiomas de Euclides. Así pues, y de acuerdo con los conocimientos de la ciencia actual, debemos concluir que Platón estaba equivocado. *Podemos* percibir círculos perfectos en la realidad física (es decir, en la realidad virtual), pero, en cambio, nunca podremos percibirlos en el dominio de las formas porque, aun aceptando que exista, carecemos de toda percepción acerca de él.

Por cierto, la idea de Platón de que la realidad física consiste en imitaciones imperfectas de abstracciones perfectas parece hoy día innecesariamente restrictiva. Como Platón, estudiamos las abstracciones por sí mismas, pero en la ciencia posgalileana, así como en la teoría de la realidad virtual, contemplamos también las abstracciones como medios para comprender entidades *físicas* reales o artificiales, y en este contexto damos por sentado que las abstracciones son casi siempre *aproximaciones* a la verdadera situación física. Así, mientras Platón veía en los círculos terrenales trazados sobre la arena aproximaciones a los círculos auténticos, matemáticos, un físico moderno consideraría que un círculo

matemático no es más que una mala aproximación a las formas reales de las órbitas planetarias, los átomos y otros objetos físicos.

Ya que siempre podrá darse el caso de que fallen el generador de realidad virtual o su usuario, ¿es posible que una representación en realidad virtual de un círculo euclídeo llegue a alcanzar la perfección absoluta, de acuerdo con los estándares de la certeza matemática? Sí. No es que se pretenda que las matemáticas, por sí mismas, estén libres de *esa* clase de incertidumbres. Los matemáticos se pueden equivocar en sus cálculos, recordar mal axiomas, cometer errores al escribir sus trabajos, etcétera. Lo que se afirma es que, *deslices aparte*, sus conclusiones son infalibles. De modo parecido, un generador de realidad virtual, si funcionase correctamente según sus especificaciones de diseño, reproduciría a la perfección un círculo euclídeo perfecto.

Una objeción parecida sería la de que nunca podremos estar seguros de que un generador de realidad virtual se comporte correctamente cuando ejecuta un determinado programa, ya que ello dependerá de su correcto funcionamiento y, en último extremo, de las leyes de la física. Puesto que no estamos seguros de conocer a la perfección estas últimas, no podemos tener la seguridad de que la máquina esté reproduciendo de manera genuina la geometría euclídea. Una vez más, nadie niega que algún fenómeno físico imprevisto —ya sea el resultado de leyes de la física desconocidas o, simplemente, de alguna enfermedad cerebral o de una tinta trucada— pueda despistar a un matemático. Pero, si las leyes de la física son, en los aspectos relevantes, como creemos, un generador de realidad virtual podrá realizar su tarea a la perfección, por más que no podamos estar seguros de ello. A este respecto, debemos evitar cuidadosamente mezclar dos cuestiones distintas: si *podemos saber* que un generador de realidad virtual reproduce un círculo perfecto y si lo hace *realmente.* Nunca lo podremos saber con seguridad, pero ello no disminuye ni un ápice la perfección con que una máquina puede reproducirlo. Volveré sobre esta distinción crucial —entre el conocimiento perfecto (certeza) de una entidad y el hecho de que ésta sea «perfecta»— en breve.

Supongamos que modificamos deliberadamente el programa de geometría euclídea de modo que un generador de realidad virtual represente los círculos aceptablemente, pero no con absoluta perfección.

¿Seríamos entonces incapaces de inferir *algo* sobre los círculos perfectos a partir de la experiencia de esa reproducción imperfecta? Ello dependería por completo de nuestro conocimiento de los aspectos del programa que hubiesen sido alterados. Si lo supiéramos, podríamos averiguar con certeza (despistes aparte) qué aspectos de las experiencias que tuviésemos dentro de la máquina representarían fielmente círculos perfectos y cuáles no. En este caso, el conocimiento adquirido sería tan fiable como el obtenido mediante la utilización del programa correcto.

Cuando *imaginamos* círculos, representamos exactamente esa clase de realidad virtual en nuestros cerebros. La razón por la que esta manera de pensar acerca de los círculos perfectos no es inútil, es que somos capaces de elaborar teorías precisas sobre qué propiedades comparten nuestros círculos imaginarios con los perfectos y cuáles no.

Utilizando una representación en realidad virtual perfecta, podríamos tener la experiencia de ver seis círculos idénticos tangentes a otro círculo idéntico, en el mismo plano y sin superponerse. Esta experiencia, en esas circunstancias, sería el equivalente de una demostración rigurosa de que dicha conformación es posible, puesto que las propiedades geométricas de las formas reproducidas serían absolutamente idénticas a las de las formas abstractas. Sin embargo, esta especie de interacción «forzada» con formas perfectas no puede aportar *todos* los conocimientos de la geometría euclídea. La mayor parte de sus fenómenos interesantes no se refieren a una conformación geométrica determinada, sino a infinitas clases de conformaciones. Por ejemplo, la suma de los ángulos de cualquier triángulo euclídeo es siempre de 180°. Podemos medir cierto número de triángulos con absoluta exactitud mediante la realidad virtual, pero ni siquiera ésta nos permitirá medir nunca todos los triángulos posibles, por lo que no podremos verificar el teorema.

¿Cómo verificarlo? Demostrándolo. Una demostración se define tradicionalmente como una secuencia de enunciados que satisfacen reglas autoevidentes de inferencia, pero ¿cuál es el equivalente físico del *proceso* de «demostración»? Para demostrar un enunciado sobre infinitos triángulos de una sola vez, examinamos determinados objetos físicos —en este caso, símbolos— que comparten propiedades comunes con clases enteras de triángulos. Por ejemplo, cuando, en circunstancias apropiadas, observamos los símbolos que dicen «Δ ABC ≡ DEF» (es

decir, «el triángulo ABC es congruente con el triángulo DEF»), concluimos que toda una clase de triángulos que hemos definido de un modo determinado, tendrá siempre la misma forma que los correspondientes triángulos de otra clase, que hemos definido de un modo distinto. Las «circunstancias apropiadas» que otorgan a esta conclusión la categoría de demostración son, en términos físicos, que los símbolos aparecen en una página por debajo de otros símbolos (algunos de los cuales representan axiomas de la geometría euclídea) y que la disposición con que aparecen esos símbolos se adapta a ciertas reglas, a saber, las reglas de inferencia.

Pero ¿qué reglas de inferencia deberíamos utilizar? Es lo mismo que preguntar cómo deberíamos programar un generador de realidad virtual para que reprodujese el mundo de la geometría euclídea. La respuesta es que debemos utilizar unas reglas de inferencia que, de la mejor manera que nos sea posible, hagan que nuestros símbolos se comporten, en aspectos relevantes, como las entidades abstractas que representan. ¿Cómo podemos estar seguros de que lo harán? No podemos. Supongamos que algunos críticos pusieran reparos a nuestras reglas de inferencia porque pensaran que los símbolos que utilizamos se comportarán de modo distinto al de las entidades abstractas. No podemos apelar a la autoridad de Aristóteles o Platón, ni demostrar que nuestras reglas de inferencia sean infalibles (muy al contrario que el teorema de Gödel, esto nos llevaría a una regresión infinita, ya que primero deberíamos demostrar que el método de demostración empleado es en sí mismo válido). Tampoco podemos contestarles arrogantemente a esos críticos que algo debe fallar en su intuición, porque la *nuestra* nos dice que los símbolos imitarán perfectamente a las entidades abstractas. Todo lo que podemos hacer es explicar. Debemos explicar por qué creemos que, en las circunstancias adecuadas, los símbolos se comportarán del modo deseado, según las reglas que proponemos. Y los críticos podrán explicar por qué se inclinan por una teoría contraria. Un desacuerdo acerca de dos teorías rivales es, en parte, un desacuerdo acerca del comportamiento observable de objetos físicos. Tales desacuerdos pueden ser tratados con los métodos normales de la ciencia. A veces pueden ser fácilmente resueltos, y a veces, no. Otra causa para semejantes diferencias de opinión, podría ser un conflicto conceptual respecto de la naturaleza de las propias entidades abstractas. De nuevo, se trata de una rivalidad

entre explicaciones, en esta ocasión acerca de abstracciones más que acerca de objetos físicos. O bien llegaríamos a un entendimiento común con nuestros críticos, o bien convendríamos en que estábamos discutiendo acerca de objetos abstractos distintos, o bien no nos pondríamos de acuerdo en nada. No existen garantías. Queda, pues, claro que, al contrario de la creencia tradicional, no es cierto que las disputas matemáticas puedan ser resueltas siempre por medios basados puramente en el debate.

A primera vista, una demostración simbólica convencional parece tener un carácter bien distinto de la demostración «forzada» realizada mediante la realidad virtual. Pero ahora vemos que están relacionadas, en la misma medida en que el cálculo lo está con los experimentos físicos. Todo experimento físico puede ser considerado un cálcu

lo, y todo cálculo es un experimento físico. En ambas clases de demostración se manipulan entidades físicas (mediante la realidad virtual o no) de acuerdo con unas determinadas reglas. En ambos casos las entidades físicas representan a las entidades abstractas que nos interesan. En ambos casos, finalmente, la fiabilidad de la demostración se basa en la verdad de la teoría de que las entidades físicas y las abstractas comparten realmente las adecuadas propiedades.

De lo que acabamos de decir, se sigue que la demostración es un *proceso* físico. De hecho, una demostración es una clase de cálculo. «Demostrar» una proposición significa realizar un cálculo que, hecho correctamente, establecerá que la proposición es cierta. Al utilizar la palabra «demostración» para indicar un *objeto* -por ejemplo, un texto escrito con tinta sobre un papel—, queremos decir que el objeto puede ser utilizado como programa para recrear un cálculo de la clase adecuada.

Y se sigue también que ni los teoremas matemáticos, ni los procesos de demostración matemática, ni la experiencia de la intuición matemática, confieren certeza alguna. Nada la confiere. Nuestro conocimiento matemático puede, al igual que nuestro conocimiento científico, ser profundo y amplio, así como sutil y maravillosamente explicativo, y ser aceptado, además, sin la menor controversia, pero nunca podrá ser seguro. Nadie puede garantizar que un buen día no se descubra que una demostración previamente considerada válida contenía un grave error, inadvertido hasta entonces, consecuencia de una asunción

«autoevidente» y nunca cuestionada antes sobre el mundo físico, o el mundo abstracto, o el modo en que están relacionadas determinadas entidades físicas y abstractas.

Una errónea asunción «autoevidente» de esta índole provocó que la propia geometría fuera incorrectamente clasificada como una rama las matemáticas durante más de dos milenios, desde aproximadamente el año 300 a.C., en que Euclides escribió sus *Elementos*, hasta el siglo XIX (y, de hecho, en muchos diccionarios y libros de texto actuales). La geometría euclídea formaba parte de la intuición de todo matemático. Con el tiempo, algunos de ellos empezaron a dudar de que cierto axioma de Euclides fuera «autoevidente» (en concreto, el llamado «axioma paralelo»). No dudaban, al menos al principio, de que el axioma fuese cierto. Se cree que el gran matemático alemán Cari Friedrich Gauss fue el primero en someterlo a comprobación. El axioma paralelo interviene en la demostración de que los ángulos de un triángulo suman 180°. Según la leyenda, en el mayor de los secretos (por temor al ridículo), Gauss situó a ayudantes dotados de linternas y teodolitos en las cimas de tres colinas que constituían los vértices del mayor triángulo que podía medir con relativa facilidad. No detectó desviación alguna de las predicciones de Euclides, pero ahora sabemos que ello se debió a que sus instrumentos no eran lo bastante sensibles. (La cercanía de la superficie terrestre es un lugar muy tranquilo, geométricamente hablando.) La teoría general de la relatividad de Einstein incluye una nueva teoría de la geometría que contradice a Euclides y ha sido corroborada por la experimentación. Los ángulos de un triángulo real *no* suman necesariamente 180°: el verdadero total depende del campo gravitatorio en el interior del triángulo.

Una clasificación incorrecta muy parecida es consecuencia del error fundamental que los matemáticos han venido cometiendo desde la antigüedad acerca de la naturaleza intrínseca de su disciplina, a saber, que el conocimiento matemático es más seguro que cualquier otro. Tras haber cometido este error, no quedaba otra elección que considerar a la teoría de la demostración parte de las matemáticas, puesto que un teorema matemático no podría ser cierto si la teoría que justifica su método de demostración fuese de por sí falsa. Pero, como acabamos de ver, la teoría de la demostración no es una rama de las matemáticas, es una ciencia. Las demostraciones no son abstractas. No es posible demostrar nada de manera abstracta, del mismo modo que no es posible

calcular nada de manera abstracta. Podemos, por supuesto, definir una clase de entidades abstractas y denominarlas «demostraciones», pero estas «demostraciones» no pueden verificar enunciados matemáticos, puesto que nadie las ve. No podrán persuadir a nadie de la autenticidad de una proposición, del mismo modo que un generador de realidad virtual que no exista físicamente no puede persuadir a nadie de que se encuentra en un entorno distinto, o un ordenador abstracto no nos puede factorizar ningún número. Una «teoría de las demostraciones» matemática no

tendría ningún efecto sobre las verdades matemáticas que pueden ser • demostradas o no en realidad, del mismo modo que una teoría de la «calculabilidad» abstracta carecería de todo efecto sobre lo que los matemáticos (o cualquiera) puedan calcular o no en realidad, a menos de que exista una razón empírica independiente para creer que los «cálculos» abstractos de la teoría se parecen a los reales. Los cálculos, incluyendo aquellos específicos que satisfacen los requisitos necesarios para ser considerados demostraciones, son procesos físicos. La teoría de la demostración trata de cómo asegurarse de que dichos procesos imiten correctamente a las entidades abstractas que se supone que deben imitar.

Los teoremas de Gödel han sido ensalzados como «los primeros teoremas nuevos de la lógica pura durante dos mil años», pero esto no es cierto. Sus teoremas tratan de lo que puede ser demostrado o no, y la demostración es un proceso físico. Ninguna faceta de la teoría de la demostración depende únicamente de la lógica. La nueva manera en que Gödel consiguió demostrar aserciones de carácter general acerca de demostraciones se basa en determinadas asunciones acerca de los procesos que pueden representar o no un hecho abstracto de un modo que el observador puede percibir y que resulte convincente para él. Gödel resumió tales asunciones en su justificación explícita y tácita de sus resultados. Éstos no estaban autoevidente-mente justificados porque fueran «pura lógica», sino porque los matemáticos encontraron que dichas asunciones eran autoevidentes.

Una de las asunciones de Gödel era la clásica de que una demostración sólo puede tener un número finito de pasos. La justificación intuitiva de esta asunción es que somos seres finitos y nunca podríamos llegar a captar un número literalmente infinito de

aserciones. Esta asunción, por cierto, provocó inquietud entre muchos matemáticos cuando, en 1976, Kenneth Appel y Wolfgang Haken utilizaron un ordenador para demostrar la famosa «conjetura de los cuatro colores» (la de que utilizando sólo cuatro colores, puede ser coloreado cualquier mapa dibujado sobre un plano sin que dos zonas adyacentes tengan nunca el mismo color). El programa requirió cientos de horas de cálculos realizados por el ordenador, lo que significaba que los pasos de la demostración, puestos por escrito, no hubiesen podido ser leídos —y mucho menos reconocidos como autoevidentes— por un ser humano en varias vidas. «¿Nos fiamos de la palabra del ordenador y damos la conjetura por demostrada?», dudaban los escépticos, maravillados, aunque nunca se les habría ocurrido catalogar todas las descargas de todas las neuronas de sus propios cerebros cuando aceptaban una demostración relativamente «sencilla».

Esa preocupación parece aún más justificada cuando la aplicamos a una presunta demostración con un número infinito de pasos. Pero ¿qué es un «paso» y qué es «infinito»? En el siglo V a.C. Zenón de Elea concluyó, sobre la base de una intuición similar, que, en una hipotética carrera entre Aquiles y una tortuga, aquél nunca podría dar alcance al animal, a condición de que éste iniciara la carrera con ventaja. Es evidente que, mientras Aquiles se dirige al punto donde se encuentra ahora la tortuga, ésta habrá avanzado algo. Cuando el atleta haya alcanzado ese punto, el animal habrá vuelto a avanzar un poco, y así *ad infinitum.* Por consiguiente, el proceso de «alcanzar» requiere que Aquiles realice un número infinito de pasos, lo cual, como ser finito que es, se supone que no podrá hacer. Pero lo que Aquiles es capaz de hacer no puede ser descubierto por la lógica pura. Se basa única y exclusivamente en lo que las leyes de la física que gobiernan el movimiento digan que puede hacer, y si dicen que adelantará a la tortuga, sin duda, lo hará. De acuerdo con las leyes de la física clásica, dar alcance a la tortuga requeriría un número infinito de pasos del tipo «desplazarse a su posición actual», y, en este sentido, se trataría de una operación infinita desde un punto de vista calculatorio. Del mismo modo, considerada como una demostración de que una cantidad abstracta se vuelve mayor que otra cuando se aplica una determinada serie de operaciones, constituye una demostración con un número infinito de pasos. Pero las leyes que rigen el movimiento lo consideran un proceso físico finito, y eso es lo que cuenta.

La intuición de Gödel acerca de los pasos y el infinito capta, en la medida de nuestros conocimientos, las auténticas limitaciones físicas del proceso de demostración. La teoría cuántica requiere pasos discretos, y ninguna de las maneras conocidas en que pueden interactuar los objetos físicos permitiría la existencia previa de un número infinito de pasos para llegar a una conclusión mensurable. (Podría, sin embargo, ser posible que se llevase a cabo un número infinito de pasos en la historia total del universo, como explicaré en el capítulo 14.) La física clásica no se habría sujetado a esas intuiciones en el supuesto, imposible, de haber sido cierta. El movimiento continuo de los sistemas clásicos, por ejemplo, habría permitido un cálculo «análogo» que no procedería a pasos y tendría un repertorio sustancialmente distinto del de la máquina universal de Turing. Se conocen varios ejemplos de leyes clásicas artificiales según las cuales podría realizarse una cantidad infinita de cálculos (es decir, infinita para los estándares de la máquina de Turing o el ordenador cuántico) mediante métodos físicamente finitos. Por supuesto, la física clásica es incompatible con los resultados de muchísimos experimentos, de modo que resulta más bien artificial especular acerca de lo que «hubiesen podido ser» las leyes clásicas «reales» de la física, pero lo que esos ejemplos nos muestran es que, con independencia de los conocimientos que se tengan de física, no se puede *probar* que una demostración deba consistir en un número infinito de pasos. Las mismas consideraciones rigen para la intuición de que debe haber un número grande, pero finito, de reglas de inferencia y de que éstas deben ser «directamente aplicables». Ninguno de estos requerimientos tiene el menor significado en sentido abstracto: son requerimientos físicos. Hilbert, en su influyente trabajo acerca del infinito, ridiculizó despiadadamente la idea de que la exigencia de un «número finito de pasos» figure entre los requerimientos fundamentales, pero el argumento precedente demuestra que estaba equivocado: es fundamental, y ello se desprende, ni más, ni menos, de la intuición *física* del propio Hilbert y otros matemáticos.

Una, al menos, de las intuiciones de Gödel sobre la demostración ha resultado ser errónea. Afortunadamente, ello no afecta a las demostraciones de sus teoremas. La heredó intacta de la prehistoria de las matemáticas griegas y permaneció incuestionada por todas las generaciones de matemáticos hasta los años ochenta, cuando se demostró su falsedad gracias a los

avances en la teoría cuántica de la calculabilidad. Se trata de la intuición de que una demostración es una clase particular de *objeto*, a saber, una secuencia de enunciados que obedecen a reglas de inferencia. He argumentado ya que una demostración debe considerarse más bien un proceso, una clase de cálculo, que un objeto, pero esta diferenciación no tiene una importancia fundamental en la teoría clásica de la demostración, ni en la de la calculabilidad, por la siguiente razón: si podemos realizar un proceso de demostración, podremos, tan sólo con un moderado esfuerzo adicional, llevar un registro de todo lo relevante que ocurra durante dicho proceso. Este registro —un objeto físico- constituirá de por sí una demostración, en el sentido de secuencia de enunciados. Inversamente, si disponemos de tal registro, podremos consultarlo y comprobar si satisface las reglas de inferencia; si lo hace, habremos demostrado la conclusión. En otras palabras, en el caso clásico, la conversión de los procesos de demostración en objetos de demostración, y viceversa, es siempre una tarea tratable.

Consideremos ahora un cálculo matemático intratable por los ordenadores clásicos, pero que un ordenador cuántico pudiese realizar con facilidad utilizando la interferencia entre, digamos, 10^{500} universos. Para aclarar un poco más el tema, supongamos que el cálculo es tal que la respuesta (a diferencia de un resultado de factorización) no podrá ser verificada tratablemente cuando la obtengamos. El proceso de programación de un ordenador cuántico para realizar dicho cálculo, ejecutar el programa y obtener un resultado, constituye una demostración de que el cálculo matemático arroja ese resultado particular. En este caso, sin embargo, no resulta posible mantener un registro de lo ocurrido durante el proceso de demostración, puesto que en su mayor parte sucedió en otros universos, y medir el estado calculatorio equivaldría a alterar las propiedades de interferencia y, por consiguiente, a invalidar la demostración. Así pues, no es factible crear un anticuado *objeto* de demostración. Es más, no existe, ni remotamente, material suficiente en el universo, tal como lo conocemos, para hacer un objeto así, ya que habría infinitamente más pasos en la demostración que átomos en el universo conocido. Este ejemplo demuestra que, a causa de la posibilidad del cálculo cuántico, las dos nociones de demostración no son equivalentes. La intuición de que una demostración es un objeto no engloba todas las formas en que un enunciado

matemático puede ser demostrado en realidad.

Una vez más, comprobamos lo inadecuado del método matemático tradicional, consistente en obtener una certeza mediante el intento de erradicar toda posible fuente de ambigüedad o error de nuestras intuiciones hasta que tan sólo quede la verdad autoevidente. Esto es lo que hizo Gödel y lo que hicieron Church, Post y, particularmente, Turing, al tratar de intuir sus modelos universales de calculabilidad. Turing esperaba que su modelo abstracto de cinta de papel resultara tan simple, tan transparente y bien definido que no necesitara apoyarse en ninguna asunción sobre física que pudiera ser concebiblemente falseada, y, por consiguiente, que dicho modelo se convirtiera en el fundamento de una teoría abstracta de la calculabilidad, independiente de la física que le sirviera de base. «Creía», como dijo Feynman, «que comprendía el papel.» Pero se equivocaba. El papel real, cuantomecánico, es muy distinto de la materia abstracta que utiliza la máquina de Turing. Ésta es enteramente clásica, y no permite, por consiguiente, la posibilidad de que el papel pueda llevar escritos distintos símbolos en diferentes universos capaces de interferir entre sí. Por supuesto, no es factible detectar la interferencia entre diferentes estados de una cinta de papel, pero el punto importante es que la intuición de Turing, a causa de que incluía falsas asunciones tomadas de la física clásica, provocó que hiciese caso omiso de algunas propiedades *calculatorias* de su hipotética máquina, precisamente aquellas que le interesaba mantener. Por esta razón el modelo de cálculo resultante era incompleto.

De todos modos, resulta muy natural que los matemáticos hayan cometido, a lo largo del tiempo, diversos errores al enfrentarse a los problemas planteados por la demostración y la certeza. El presente estudio debería convencernos de que la concepción actual tampoco será eterna. Pero la complacencia con que los matemáticos han caído en esos errores, y su incapacidad para reconocer ni siquiera la posibilidad de equivocarse en las materias mencionadas están, en mi opinión, relacionadas con la antigua y extendida confusión entre los *métodos* matemáticos y su *materia.* Permítaseme explicarme. A diferencia de las relaciones entre entidades físicas, las que vinculan a las entidades abstractas son independientes de cualquier hecho contingente y de las leyes de la física. Están absoluta y objetivamente determinadas por las propiedades autónomas de las propias entidades abstractas. Las

matemáticas, el estudio de estas relaciones y propiedades, es, por lo tanto, el estudio de *verdades absolutamente necesarias.* En otras palabras, las verdades que *estudian* las matemáticas son absolutamente ciertas. Pero ello no significa que nuestro conocimiento de estas verdades sea en sí mismo cierto, ni que los métodos empleados en dicho estudio confieran por fuerza veracidad a sus conclusiones. Después de todo, las matemáticas estudian también equívocos y paradojas, sin que ello signifique que las conclusiones de sus estudios deban ser a su vez equívocas o paradójicas.

La verdad necesaria constituye, simplemente, la *materia de estudio* para las matemáticas y no la recompensa que obtenemos mediante su práctica. El objetivo de las matemáticas no es, ni puede serlo, la certeza matemática, ni tan sólo la verdad matemática, cierta o no. Es, y debe serlo, la explicación matemática.

¿Por qué, entonces, funcionan tan bien las matemáticas? ¿Cómo es que conducen a conclusiones que, si bien no son absolutamente ciertas, pueden ser aceptadas y aplicadas sin problemas, al menos, durante milenios? En última instancia, la razón es que *parte* de nuestro conocimiento del mundo físico es igual de fiable e incontrovertible, y al comprender este mundo lo suficientemente bien, podemos también comprender qué objetos físicos tienen propiedades comunes con objetos abstractos, y cuáles son éstos. Así pues, en principio, la fiabilidad de nuestro conocimiento matemático está sujeta a lo que sepamos de la realidad física. La validez de toda demostración matemática depende por completo de lo acertada que sea nuestra comprensión de las reglas por las que se rige el comportamiento de ciertos objetos físicos, sean éstos generadores de realidad virtual, papel y tinta, o nuestros propios cerebros.

Así pues, la intuición matemática es una variedad de la intuición física. Ésta es una serie de reglas empíricas, algunas tal vez innatas, muchas elaboradas durante la infancia, acerca de cómo se comporta el mundo. Tenemos, por ejemplo, intuiciones de que hay objetos físicos, y de que éstos poseen atributos tales como forma, color, peso y posición en el espacio, algunos de los cuales existen incluso cuando los objetos no son observados por nadie. Otra es la de que hay una variable —el tiempo— con respecto a la cual dichos atributos cambian, aunque sin perder su identidad mientras transcurre. Otra es que los objetos interactúan,

y que ello puede provocar algunos cambios en sus atributos. La intuición matemática se ocupa del modo en que el mundo físico es capaz de mostrar las propiedades de las entidades abstractas. Una de estas intuiciones es la de que hay una ley abstracta o, al menos, una explicación que constituye la base del comportamiento de los objetos. La intuición de que el espacio admite superficies cerradas que separan un «dentro» de un «fuera» puede ser aquilatada hasta llegar a la intuición matemática de *conjunto*, que lo divide todo en miembros y no miembros de éste. Pero ulteriores aquilataciones, ahora realizadas por matemáticos (empezando con la refutación por Russell de la teoría de conjuntos de Frege), han demostrado que esa intuición deja de ser precisa cuando los conjuntos en cuestión contienen «demasiados» miembros (un grado excesivo de infinidad de miembros).

Aunque una intuición física o matemática fuese innata, ello no le conferiría ninguna autoridad particular. La intuición innata no puede ser tomada por un sucedáneo de los «recuerdos» de Platón del mundo de las formas, ya que es una observación bien conocida que muchas de las intuiciones que han ido elaborando los seres humanos como consecuencia de accidentes durante la evolución son, simplemente, erróneas. Por ejemplo, el ojo humano y los restantes órganos de la visión llevan incorporada, implícitamente, la errónea teoría de que la luz amarilla es una mezcla de luz verde y luz roja (en el sentido de que la luz amarilla nos produce exactamente la misma sensación que una mezcla de las otras dos). En realidad, las tres clases de luz tienen frecuencias distintas, y no pueden ser creadas mezclando luces de otras frecuencias. El hecho de que una mezcla de luz roja y luz verde nos parezca luz amarilla no tiene nada que ver, ni mucho menos, con las propiedades de la luz, sino que es una propiedad de nuestros ojos, resultado de un compromiso de diseño ocurrido en algún momento de la evolución de nuestros más lejanos antepasados. Quizás sea posible (aunque, personalmente, no lo creo) que la geometría euclídea o la lógica aristotélica estén preinstaladas de algún modo en la estructura de nuestros cerebros, como creía el filósofo Immanuel Kant, pero ello no implicaría de un modo lógico que fuesen ciertas. Incluso en el, aún más inverosímil, caso de que tuviésemos intuiciones innatas de las que fuésemos constitucionalmente incapaces de librarnos, dichas intuiciones no tendrían por qué ser ciertas.

La estructura de la realidad tiene, pues, una trama mucho

más unificada de lo que hubiera resultado posible de haber sido el conocimiento matemático verificable con certeza y, por consiguiente, jerárquico, como se ha creído tradicionalmente. Las entidades matemáticas forman parte de la estructura de la realidad porque son complejas y autónomas. La clase de realidad que conforman es, en algunos aspectos, parecida al reino de las abstracciones contemplado por Platón

o por Penrose: si bien son intangibles por definición, existen de manera objetiva y tienen propiedades que son independientes de las leyes de la física. Sin embargo, es ésta la que nos permite adquirir conocimiento de ese reino. Pero la física nos impone también severas restricciones. Mientras que en la realidad física todo es comprensible, las verdades matemáticas comprensibles son tan sólo las de la minoría infinitesimal que se corresponde exactamente con alguna verdad física, como el hecho de que, cuando determinados símbolos trazados con tinta sobre un papel son manipulados de cierto modo, aparecen otros símbolos. Es decir, son verdades que pueden ser representadas mediante la realidad virtual. No tenemos más remedio que asumir que las entidades matemáticas incomprensibles son también reales, puesto que aparecen ineluctablemente en nuestras explicaciones de las entidades comprensibles.

Existen objetos físicos —como los dedos, los ordenadores y los cerebros— cuyo comportamiento puede personificar el de ciertos objetos abstractos. De este modo, la estructura de la realidad física nos ofrece una ventana abierta al mundo de la abstracción. Es una ventana muy estrecha, que nos proporciona tan sólo una limitada perspectiva. Algunas de las estructuras que vemos en el exterior, como los números naturales o las reglas de inferencia de la lógica clásica, parecen ser importantes o «fundamentales» para el mundo abstracto, del mismo modo que las más complejas leyes de la naturaleza lo son para el mundo físico, pero podría tratarse de una apariencia engañosa. Lo que realmente vemos es tan sólo que algunas estructuras abstractas son fundamentales *para nuestra comprensión* de las abstracciones. No hay motivo para suponer que esas estructuras sean significativas, objetivamente, en el mundo abstracto. Se trata, sólo, de que algunas de estas entidades abstractas están más próximas a nuestra pequeña ventana y se ven mejor desde ella que otras.

TERMINOLOGÍA

Matemáticas. Estudio de las verdades absolutamente necesarias.

Demostración. Modo de establecer la verdad de las proposiciones matemáticas.

Definición tradicional. Secuencia de enunciados que se inicia con algunas premisas y termina con la conclusión deseada, y que satisface determinadas «reglas de inferencia».

Definición mejorada. Cálculo que simula las propiedades de alguna entidad abstracta, y cuyo resultado establece que dicha entidad tiene una determinada propiedad.

Intuición matemática. *Tradicionalmente: Fuente de justificación definitiva y autoevidente para el razonamiento matemático. Actualmente:* Conjunto de teorías (conscientes o inconscientes) sobre el comportamiento de determinados objetos físicos que se asemeja al de entidades abstractas interesantes.

Intuicionismo. Doctrina según la cual ningún razonamiento sobre entidades abstractas es fiable, a menos que se base en la intuición directa y autoevidente. Constituye la versión matemática del solipsismo.

Décimo problema de Hilbert. Establecer de una vez por todas la certeza de los métodos matemáticos mediante el hallazgo de un conjunto de reglas de inferencia suficiente para todas las demostraciones válidas, y posterior demostración de que esas reglas son consistentes con sus propios estándares.

Teorema de la incompletitud de Gödel. Demostración de que el décimo problema de Hilbert es ínsoluble. Enuncia que para todo conjunto de reglas de inferencia existen demostraciones válidas no consideradas como tales por esas reglas.

SUMARIO

Desde un punto de vista objetivo, hay entidades abstractas que son complejas y autónomas y forman parte de la estructura de la realidad, y hay, asimismo, verdades lógicamente necesarias acerca de ellas, entre las cuales se encuentran las matemáticas en cuanto disciplina. No obstante, esas verdades no pueden ser conocidas con certeza. Las demostraciones no proporcionan certeza sobre sus conclusiones. La validez de una determinada forma de demostración depende de la verdad de nuestras teorías sobre el comportamiento de los objetos con los que realizamos dicha demostración. Por consiguiente, el conocimiento matemático es intrínsecamente derivativo y está supeditado por completo a nuestros conocimientos de física. Las verdades matemáticas comprensibles son tan sólo las de la minoría infinitesimal que puede ser representada mediante la realidad virtual. Sin embargo, las entidades matemáticas abstractas (por ejemplo, los entornos cantgotu) existen también, puesto que aparecen ineluctablemente en nuestras explicaciones de las entidades comprensibles.

He afirmado que el cálculo ha sido siempre un concepto cuántico, ya que la física clásica es incompatible con las intuiciones que forman la base de la teoría clásica de la calculabilidad. Lo mismo es cierto para el tiempo. Milenios antes del descubrimiento de la teoría cuántica, el tiempo ya era el concepto cuántico fundamental.

11. EL TIEMPO, CONCEPTO CUÁNTICO FUNDAMENTAL

Como las olas en su camino hacia la orilla pedregosa,
raudos nuestros minutos hacia su fin tienden;
reemplazando cada uno al que lo precedió,
en esfuerzo encadenado, todos por avanzar contienden.

Soneto LX, WILLIAM SHAKESPEARE

A pesar de ser uno de los atributos más familiares del mundo físico, el tiempo goza de la reputación de ser profundamente misterioso. Este misterio forma parte del mismo concepto del tiempo que damos por sentado. San Agustín dijo, por ejemplo: «¿Qué es, pues, el tiempo? Si alguien me lo pregunta, lo sé; cuando deseo explicárselo a quien me interroga, no puedo» *(Confesiones).*

Pocas personas consideran misteriosa la distancia, pero todo el mundo sabe que el tiempo lo es. Y el misterio del tiempo nace de su atributo más básico y lógico, a saber, que el momento presente que denominamos «ahora» no es fijo, sino que se mueve continuamente en la dirección del futuro. Denominamos a ese movimiento *transcurrir* el tiempo.

Veremos que, en realidad, el tiempo no transcurre. Pero la idea de que lo hace se basa en el más puro sentido común. Es algo que damos por sentado, que está asumido en la propia estructura de nuestro lenguaje. En *A Comprehensive Grammar of the English Language*, Randolph Quirk y sus colaboradores explican el concepto del tiempo con ayuda del diagrama que muestra la figura 11.1. Cada punto de la línea representa un momento particular, fijo. El triángulo (∇) indica dónde se halla situado en la línea «el punto en movimiento continuo, el momento presente». Se supone que se mueve de izquierda a derecha. Algunas personas, como Shakespeare en el soneto transcrito, consideran que los acontecimientos particulares están «fijos» y la línea pasa por ellos al moverse (de derecha a izquierda en la figura 11.1), de modo que los momentos del futuro pasan por el momento presente para convertirse en momentos del pasado.

¿Qué significa eso de que «el tiempo puede ser considerado una línea»? Significa que, del mismo modo que una línea puede ser vista como una secuencia de puntos en distinta posición, cualquier objeto que cambia o se mueve puede ser considerado una secuencia de «instantáneas», versiones fijas de sí mismo, una para cada momento. Decir que cada punto de la línea representa un determinado momento equivale a decir que podemos imaginar todas las instantáneas dispuestas correlativamente sobre la línea, como en la figura 11.2. Algunas de ellas muestran la flecha en rotación tal como estaba en el pasado, y otras tal como estará en el futuro, mientras que una de ellas —la que señala el triángulo en movimiento— nos muestra la flecha tal como está ahora, si bien un momento después esta versión particular formará ya parte del pasado, puesto que el triángulo se habrá movido hacia adelante. Las versiones instantáneas de un objeto *son* colectivamente ese objeto en movimiento, en el mismo sentido en que las imágenes fijas proyectadas en una secuencia sobre una pantalla *son* colectivamente una película animada. Ninguna de ellas, individualmente, cambia jamás. El cambio consiste en el hecho de ser señaladas («iluminadas») en secuencia por el triángulo en movimiento (el «proyector») de modo que, una tras otra, les llega el turno de estar en el presente.

—Hoy día los lingüistas rehuyen hacer juicios de valor sobre el modo en que es utilizado el lenguaje y se limitan a registrarlo, analizarlo y comprenderlo. Por consiguiente, Quirk y sus colaboradores no deben ser criticados por la calidad de la teoría del tiempo que exponen. No defienden que sea una buena teoría, tan sólo afirman —y creo que con razón— que es *nuestra* teoría. Pero, por desgracia, no es una buena teoría. Para decirlo con franqueza, la única razón de que la teoría del sentido común del tiempo sea intrínsecamente misteriosa es que, intrínsecamente carece de sentido. No se trata sólo de que los hechos demuestren que es inexacta. Veremos que, aun en sus propios términos, carece, realmente, de sentido.

Ello quizás resulte sorprendente. Nos hemos acostumbrado a modificar nuestro sentido común para adaptarlo a los descubrimientos científicos. El sentido común resulta ser a menudo

erróneo, incluso terriblemente erróneo, pero no es habitual que *carezca de sentido* en cuestiones de experiencia cotidiana. Sin embargo, eso es lo que ocurre en esta ocasión.

Consideremos de nuevo la figura 11.2, que ilustra el movimiento de dos entidades. Una de ellas es una flecha en rotación mostrada como una secuencia de instantáneas. La otra es el «momento presente» en movimiento, que se desplaza por la figura de izquierda a derecha. Pero el movimiento del momento presente no se muestra en la figura como una secuencia de instantáneas. En vez de ello, se selecciona un momento en particular mediante el triángulo, además de destacarlo con líneas más gruesas y la leyenda «(ahora)». Así pues, si bien, según la leyenda, «ahora» se mueve por la figura, de él sólo se muestra una instantánea, correspondiente a un momento particular.

¿Por qué? Después de todo, el objetivo de la figura es mostrarnos lo que sucede en un período extenso y no en un momento particular. Si lo que queríamos era que la figura mostrase tan sólo un momento, no teníamos por qué habernos molestado en mostrar más que una instantánea de la flecha en rotación. Se supone que la figura ilustra la teoría de sentido común de que cualquier objeto que cambia o se mueve es una secuencia de instantáneas, una para cada momento. Así pues, si el triángulo está en movimiento, ¿por qué no mostramos también su correspondiente serie de instantáneas? La única instantánea que vemos del triángulo debe ser tan sólo una de las muchas que habría, si ésta fuese una verdadera descripción de cómo funciona el tiempo. De hecho, la figura resulta altamente engañosa tal como se presenta: muestra que el triángulo *no* está en movimiento, sino que empieza a existir en un momento particular e inmediatamente llega a su fin. De ser así, ello convertiría al «ahora» en un momento *fijo.* De nada sirve haber añadido la leyenda «Movimiento del momento presente» y una flecha discontinua que indica que el triángulo se mueve hacia la derecha.

Lo que la figura en sí misma muestra, al igual que el diagrama de Quirk y sus colaboradores (figura 11.1), es que el triángulo jamás alcanza otro momento que el señalado.

Como mucho, podríamos decir que la figura 11.2 es una representación híbrida, que ilustra impropiamente el movimiento de dos formas diferentes: la flecha en rotación ilustra la teoría del tiempo según el sentido común, pero el triángulo, aunque *dice* que el momento presente se mueve, no lo muestra en movimiento. ¿Cómo podríamos modificar la figura de modo que ilustrara la teoría del tiempo de sentido común en relación tanto con el movimiento del tiempo presente como con el de la flecha? Incluyendo más instantáneas del triángulo, una por momento, cada una de las cuales indicaría dónde se encuentra «ahora» en ese momento. ¿Y dónde se encuentra? Obviamente, en cada momento, «ahora» es ese momento. A medianoche, por ejemplo, el triángulo debe apuntar a la instantánea de la flecha tomada a medianoche; a la una, deberá hacerlo hacia la instantánea correspondiente a dicha hora, y así sucesivamente. Por consiguiente, la representación resultante debería parecerse a la figura 11.3.

Esta imagen modificada ilustra de manera satisfactoria el *movimiento*, pero nos deja con un concepto del tiempo notablemente reducido. La idea de sentido común de que un objeto en movimiento es una secuencia de versiones instantáneas de sí mismo permanece, pero la otra idea de sentido común —la del transcurso del tiempo— ha desaparecido. La nueva representación carece de un «punto en continuo movimiento, el momento presente», que se desplace, uno tras otro por todos los momentos fijos. No hay en ella proceso alguno por el que un momento fijo se inicie en el futuro, se convierta en presente y quede después relegado al pasado. La multiplicidad de triángulos y «(ahora)» ya no distingue a un momento de los demás, por lo que resultan superfluos. La figura ilustraría el movimiento de la flecha en rotación igualmente bien sin ellos.

Así pues, el «momento presente» sólo existe subjetivamente. Desde el punto de vista de un observador en un momento particular, éste es, sin duda, singular, y puede ser considerado un «ahora» único por dicho observador, del mismo modo que cualquier posición en el espacio puede ser designada como un «aquí» único desde la perspectiva del observador situado en ella. Pero, objetivamente, ningún momento tiene el privilegio de ser

más «ahora» que los demás, al igual que ninguna posición tiene el privilegio de ser más «aquí» que otra. El «aquí» subjetivo se moverá por el espacio con el observador. ¿Se mueve también el «ahora» subjetivo a lo largo del tiempo? ¿Son las figuras 11.1 y 11.2 correctas, después de todo, en el sentido de ilustrar el tiempo desde el punto de vista de un observador en un momento particular? Ciertamente, no. Incluso desde un punto de vista subjetivo, «ahora» *no se mueve a lo largo del tiempo.* Se dice a menudo que el presente *parece* moverse hacia adelante en el tiempo porque el presente se define únicamente en relación con nuestra conciencia, y ésta se mueve hacia adelante a lo largo de los momentos. Pero, en realidad, nuestra conciencia no funciona, ni podría funcionar, así. Cuando decimos que nuestra conciencia «parece» pasar de un momento al próximo, simplemente parafraseamos la teoría de sentido común del transcurso del tiempo. Pero no tiene más sentido pensar que un único «momento del que somos conscientes» se desplaza de un momento a otro que pensar que lo hace un único momento presente, o cualquier cosa que se le parezca. *Nada* puede pasar de un momento a otro. Existir en un momento particular significa existir en él para siempre. Nuestra conciencia existe en *todos* nuestros momentos (de vigilia).

Sin duda, diferentes instantáneas del observador perciben como «ahora» distintos momentos, pero ello no significa que la conciencia del observador —o cualquier otra entidad que cambie o esté en movimiento— se mueva a lo largo del tiempo, como se supone que hace el momento presente. Las diversas instantáneas del observador no están por turnos en el presente. Y tampoco son conscientes de su presente por turnos. Todas ellas son conscientes y se encuentran, desde un punto de vista subjetivo, en el presente. Objetivamente hablando, no hay presente.

No experimentamos el transcurso o el paso del tiempo. Lo que experimentamos son las diferencias entre nuestras percepciones presentes y nuestros recuerdos presentes de las

percepciones pasadas. Interpretamos —correctamente— esas diferencias como la evidencia de que el universo cambia con el tiempo. También las interpretamos —incorrectamente— como la evidencia de que nuestra conciencia, o el presente, o lo que sea, se mueve a lo largo del tiempo.

Si el presente en movimiento dejara caprichosamente de desplazarse durante un par de días, para reanudar después su marcha a una velocidad diez veces superior, ¿de qué seríamos conscientes? De nada especial. O, mejor dicho, la pregunta carece de sentido. Nada hay en ese presente capaz de moverse, detenerse o transcurrir, del mismo modo que nada hay que podamos denominar significativamente «velocidad» del tiempo. Se supone que todo lo que existe en el tiempo toma la forma de instantáneas invariables, dispuestas a lo largo de la línea del tiempo. Aquí caben las experiencias conscientes de todos los observadores, incluyendo su errónea intuición de que el tiempo «transcurre». Puede que imaginen un «presente en movimiento» que viaja a lo largo de esa línea, se detiene y reanuda su marcha, o incluso que retrocede o cesa por completo de existir, pero que lo imaginen no hará que suceda. Nada puede moverse a lo largo de esa línea. El tiempo no puede transcurrir.

La idea del transcurso del tiempo presupone, en realidad, la existencia de una segunda clase de tiempo, fuera del tiempo como semencia de momentos de sentido común. Si «ahora» se moviese realmente de un momento a otro, debería hacerlo con respecto a ese tiempo *externo.* Pero tomarse esto en serio conduciría a una regresión infinita, puesto que entonces deberíamos imaginar el tiempo externo, a su vez, como una sucesión de momentos, con su propio «momento presente» que se movería con respecto a un tiempo aún más externo, y así sucesivamente. En cada nivel, el transcurso del tiempo no tendría sentido a menos que lo atribuyésemos al transcurso de un tiempo exterior, *ad infinitum.* En cada nivel nos encontraríamos pues, con un concepto que carecería de sentido. Y toda la infinita jerarquía resultante carecería igualmente de sentido.

El origen de esta clase de error es que estamos acostumbrados a que el tiempo constituya un marco exterior de toda entidad física que consideremos. Nos hemos habituado a imaginar que cualquier objeto físico puede estar potencialmente en cambio y existir, por consiguiente, como una secuencia de versiones de sí mismo en distintos momentos. Pero es que la

secuencia de los momentos, como la que muestran las figuras 11.1, 2 y 3, es de por sí una entidad excepcional. No existe dentro del marco del tiempo, *es* el marco del tiempo. Puesto que no hay tiempo fuera de ella, resulta incoherente imaginar que cambie o exista en más de una versión consecutiva. Ello hace que esas figuras resulten difíciles de comprender. La imagen que representan, como cualquier otro objeto físico, existe durante un determinado período de tiempo y consiste en múltiples versiones de sí misma, pero lo que esa imagen *describe* —a saber, la secuencia de versiones de algo— existe en una única versión. Ninguna imagen exacta del marco del tiempo puede ser cambiante o estar en movimiento, sino que debe ser estática. Hay, sin embargo, una dificultad psicológica intrínseca para asumirlo. Si bien la imagen es estática, no podemos comprenderla estáticamente. Muestra una secuencia de momentos simultáneos en una página, y para relacionarla con nuestra experiencia debemos desplazar los ojos con atención a lo largo de dicha secuencia. Podemos, por ejemplo, mirar una instantánea y considerar que representa el «ahora», y mirar un momento después la siguiente, situada a la derecha de la anterior, y pensar que representa el nuevo «ahora». Tendemos así a confundir el movimiento verdadero de nuestros ojos al mirar con atención a lo largo de la mera *imagen* con el movimiento imposible de algo a lo largo de los momentos reales. Es un error muy fácil de cometer.

Pero hay algo más en este problema que la dificultad de *ilustrar* la teoría del tiempo de sentido común. La teoría, en sí, contiene una fundamental e insuperable equivocación: no acaba de decidir si el presente es, objetivamente, un único momento o muchos, es decir, si, por ejemplo, la figura 11.1 describe un momento o varios. El sentido común exige que el presente sea un único momento para que sea posible el transcurso del tiempo, de modo que el presente pueda desplazarse a lo largo de los momentos, de pasado a futuro. Pero el sentido común exige también que el tiempo sea una secuencia de momentos y que todo cambio o movimiento consista en diferencias entre las versiones de una entidad en diferentes momentos, lo que significa que los momentos son, en sí, invariables. Así, un momento particular no puede convertirse en presente, o dejar de serlo, ya que ello conllevaría cambios. Por consiguiente, el presente no puede, objetivamente, ser un único momento.

La razón por la que nos aferramos a esos dos conceptos

incompatibles —el presente en movimiento y la secuencia de momentos invariables— es que los necesitamos, o, mejor dicho, pensamos que los necesitamos. Los invocamos constantemente en nuestra vida cotidiana, si bien nunca de manera simultánea. Cuando *describimos* sucesos, y decimos cuándo ocurren las cosas, pensamos en términos de una secuencia de momentos invariables; cuando los *explicamos* en cuanto causas y efectos unos de otros, pensamos en términos de presente en movimiento.

Por ejemplo, al decir que Faraday descubrió la inducción electromagnética «en 1831», asignamos dicho acontecimiento a un determinado conjunto de momentos. Es decir, especificamos en qué grupo de diapositivas, dentro del gran archivador de la historia del mundo, se encuentra dicho descubrimiento. No involucramos ningún transcurso del tiempo al decir *cuándo* ocurrió algo, del mismo modo que no involucramos ningún «recorrido de una distancia» si decimos *dónde* ocurrió. Pero tan pronto como explicamos *por qué* ocurrió algo invocamos el transcurso del tiempo. Si decimos que debemos en parte nuestros motores y dinamos eléctricos a Faraday, y que las repercusiones de su descubrimiento alcanzan hasta nuestros días, tenemos en mente una imagen de esas repercusiones que se inicia en 1831 y se desliza consecutivamente a lo largo de todos los momentos del resto de siglo XIX, para alcanzar después el siglo XX y provocar que surjan en él cosas como las centrales eléctricas. Si no vamos con cuidado, pensaremos en el siglo XX como «aún no afectado» inicialmente por el importante acontecimiento de 1831 y «cambiado» después por sus repercusiones, a medida que éstas avanzan en su camino hacia el siglo XXI y más allá de él. Por lo general, no obstante, somos cuidadosos y evitamos este pensamiento incoherente al no utilizar nunca de manera simultánea’las dos partes de la teoría del tiempo de sentido común. Sólo lo hacemos cuando pensamos en el tiempo propiamente dicho, y entonces, ¡nos maravillamos ante su misteriosa naturaleza! Quizás «paradoja» sería una palabra más adecuada que misterio, ya que nos enfrentamos a un conflicto descarado entre dos ideas que parecen autoevidentes. No pueden ser ciertas ambas. Veremos que, en realidad, no lo es ninguna de las dos.

Nuestras teorías físicas son, a diferencia del sentido común, coherentes, y lo primero que hicieron para conseguir esa coherencia fue abandonar la idea del transcurso del tiempo.

Ciertamente, los físicos hablan del transcurso del tiempo como cualquier otra persona. Newton, por ejemplo, escribió en sus *Principia*, obra en que sentó los principios de la mecánica y la gravitación newtonianas, lo siguiente: «El tiempo absoluto, cierto y matemático, de por sí, y por su propia naturaleza, transcurre uniformemente sin relación a nada externo.» Pero, astutamente, Newton no intentó traducir en forma matemática esta afirmación de que el tiempo transcurre, ni de derivar de ella conclusión alguna. Ninguna de las teorías físicas de Newton se refiere al transcurso del tiempo, y ninguna de las teorías subsiguientes se refiere al transcurso del tiempo o es compatible con él.

¿Qué necesidad tenía, pues, Newton de decir que «el tiempo transcurre uniformemente»? No hay nada que objetar a lo de «uniformemente»: se puede interpretar en el sentido de que las medidas de tiempo son las mismas para diferentes observadores situados en distintas posiciones y en diversos estados de movimiento. Se trata de una afirmación fundamental (que, desde Einstein, sabemos que es inexacta), y se habría podido formular, como he hecho, sin decir que el tiempo transcurre. Pienso que Newton empleó deliberadamente el lenguaje familiar de su época, sin intención de que fuera tomado en sentido literal, del mismo modo que habría podido hablar, informalmente, de la «salida» del Sol. Necesitaba convencer al lector que iniciara la lectura de su revolucionario trabajo de que no había nada nuevo o complicado en el concepto newtoniano del tiempo. Los *Principia* asignan a palabras tales como «fuerza» y «masa» significados técnicos precisos, bastante distintos de los que les asigna el sentido común, pero los números a los que se refieren como «tiempos» son, simplemente, los tiempos del lenguaje corriente que podemos encontrar en relojes y calendarios, y en los *Principia* el concepto del tiempo es el de sentido común.

Sólo que no transcurre. En la física newtoniana, tiempo y movimiento aparecen de modo muy parecido al que muestra la figura 11.3. Una pequeña diferencia es que he dibujado los momentos sucesivos separados unos de otros, pero en toda la física precuántica esto es una aproximación, porque el tiempo es un continuo. Debemos imaginar un número infinito de instantáneas infinitamente delgadas que se interpolan de manera continua entre las que he dibujado. Si cada instantánea representa todos los acontecimientos a lo largo de la totalidad del espacio que existe físicamente en un momento particular, podemos imaginar

que las diapositivas están pegadas por ambas caras formando un bloque único e invariable que contiene todo lo que acontece en el espacio y el tiempo (figura 11.4), es decir, la totalidad de la realidad física. Una inevitable limitación de esta clase de diagramas es que las instantáneas del espacio en cada momento aparecen como si fuesen bidimensionales cuando, en realidad, son tridimensionales. Cada una de ellas es el espacio en un momento particular. Así pues, tratamos al tiempo como si fuera una cuarta dimensión, análoga a las tres dimensiones del espacio de la geometría clásica. El espacio y el tiempo considerados así, en conjunto, como una entidad tetradimensional, se denominan *espacio-tiempo.*

En la física newtoniana esta interpretación geométrica tetradimensional era opcional, pero con la teoría de la relatividad de Einstein pasó a ser indispensable. Ello se debe a que, según la relatividad, distintos observadores que se mueven a diferentes velocidades no coinciden acerca de qué acontecimientos son simultáneos. Es decir, no están de acuerdo acerca de qué acontecimientos deberían aparecer en una determinada instantánea. Cada uno percibe que el espacio-tiempo es cortado en «momentos» de un modo distinto. No obstante, si todos ellos juntaran sus instantáneas al modo de la figura 11.4, obtendrían espacios-tiempo idénticos. En consecuencia, y de acuerdo con la relatividad, los «momentos» mostrados en la figura 11.4 no son características objetivas del espacio-tiempo: son, simplemente, el modo particular en que un observador percibe la simultaneidad. Un observador distinto podría dibujar los cortes de «ahora» desde un ángulo diferente. Así pues, la realidad objetiva oculta tras la figura 11.4, a saber, el espacio-tiempo y su contenido físico, podría ser expresada como en la figura 11.5.

El espacio-tiempo es denominado a veces «universo en bloque» porque incluye en su seno, en forma de bloque tetradimensional congelado, la totalidad de la realidad física pasada, presente y futura. En relación con el espacio-tiempo, nada se mueve nunca. Lo que denominamos «momentos» son determinados cortes a través del espacio-tiempo. Cuando los contenidos de dichos cortes difieren entre sí, decimos que hay cambio o movimiento en el espacio.

Como he dicho, pensamos en el transcurso del tiempo en relación con causas y efectos. Vemos a las causas como predecesoras de sus efectos; imaginamos que el presente en movimiento llega antes a las causas que a los efectos, y consideramos que éstos avanzan junto al momento presente. Filosóficamente, los procesos de causa-efecto más importantes son nuestras decisiones conscientes y las consiguientes acciones. El punto de vista del sentido común nos dice que tenemos *libre albedrío*, es decir, que nos encontramos a veces en la posición de poder influir en los acontecimientos futuros (como el movimiento de nuestros cuerpos) de alguno de diversos modos posibles y escoger, además, cuál ocurrirá, mientras que, en cambio, no tenemos ninguna posibilidad de influir sobre el pasado. (Analizaré el libre albedrío en el capítulo 13.) El pasado es fijo; el futuro está abierto. Para muchos filósofos, el transcurso del tiempo es el proceso por el que el futuro abierto se convierte, momento a momento, en el pasado fijo. Otros defienden la *posibilidad* de que ocurran sucesos alternativos en cada momento del futuro y consideran que el transcurso del tiempo es el proceso por el que, momento a momento, alguna de dichas posibilidades se convierte en *realidad* (de ese modo, y para esas personas, el futuro no existe hasta que el transcurso del tiempo lo alcanza y lo convierte en pasado). Pero si el futuro está realmente abierto (¡y lo está!),

no es en relación con el transcurso del tiempo, puesto que es algo que no existe. En la física del espacio-tiempo (en realidad, toda la física precuántica, desde Newton), el futuro no está abierto. Está *ahí*, con contenido fijo y definido, al igual que el pasado y el presente. Si un determinado momento del espacio-tiempo estuviera, en algún sentido, «abierto», lo seguiría estando al convertirse en presente y en pasado, puesto que los momentos no pueden cambiar.

Subjetivamente, podemos decir que el futuro de un determinado observador está «abierto desde el punto de vista de dicho observador», ya que uno no puede medir ni observar su propio futuro. Pero apertura no significa, en este sentido, capacidad de elección. Si usted tiene un décimo de la lotería de la semana pasada y no ha averiguado aún si le ha tocado, el resultado sigue abierto desde su punto de vista, a pesar de ser, objetivamente, fijo. Pero ni subjetiva ni objetivamente lo podrá cambiar. Ninguna causa que no lo haya afectado hasta el sorteo, lo podrá afectar ya. La teoría de sentido común del libre albedrío dice que la semana pasada, cuando aún podía elegir entre comprar o no el décimo, el futuro estaba todavía, en sentido objetivo, abierto, y usted podía realmente escoger entre dos o más opciones. Pero eso es incompatible con el espacio-tiempo. Según la física del espacio-tiempo, la apertura del futuro es una ilusión, y, por consiguiente, la causalidad y el libre albedrío no pueden ser, a su vez, más que ilusiones. Necesitamos la creencia (y nos aferramos a ella) de que el futuro puede ser influido por los acontecimientos presentes y, en especial, por nuestras elecciones, pero ello tal vez sólo sea una manera de hacer soportable el hecho de que desconocemos lo que nos espera. En realidad, no hacemos elecciones. Incluso cuando creemos que consideramos los pros y los contras de una elección, su resultado está ya ahí, en su correspondiente porción del espacio-tiempo, invariable como todo en el espacio-tiempo e insensible a nuestras consideraciones. Al parecer, esas mismas consideraciones son invariables y nos esperan ya, en sus correspondientes momentos, antes de que las hagamos.

Ser «efecto» de alguna causa significa ser afectado por dicha causa, ser cambiado por ella. Así pues, si la física del espacio-tiempo niega la realidad del transcurso del tiempo, lógicamente, tampoco puede admitir las nociones de sentido común de causa y efecto. Y es que en el universo, entendido como un bloque, nada

puede cambiar: ninguna parte del espacio-tiempo puede modificar a otra, del mismo modo que ninguna parte de un objeto *tridimensional* fijo puede cambiar a otra.

Sucede así que, en la era de la física del espacio-tiempo, todas las teorías fundamentales tenían la propiedad de que, dado todo lo que ocurra antes de un momento determinado, las leyes de la física determinan lo que sucederá en todos los momentos subsiguientes. La propiedad de que unas instantáneas sean determinadas por otras se denomina *determinismo.* En la física newtoniana, por ejemplo, si conocemos en un momento dado las posiciones y velocidades de todas las masas de un sistema aislado, como el sistema solar, podemos, en principio, calcular *(predecir)* dónde se encontrarán dichas masas en cualquier momento posterior. Podemos también, en principio, calcular *(«retrodecir»)* dónde se encontraban en cualquier momento anterior.

Las leyes de la física que diferencian una instantánea de otra son el «pegamento» que las mantiene unidas en forma de espacio-tiempo. Imaginémonos que estamos, de un modo mágico e imposible, fuera del espacio-tiempo y, por consiguiente, en nuestro propio tiempo externo, independiente del que corresponde al espacio-tiempo. Cortemos el espacio-tiempo en instantáneas de espacio correspondientes a cada momento, tal como lo percibiría un determinado observador situado en el espacio-tiempo. Barajemos ahora las instantáneas resultantes y peguémoslas en distinto orden. ¿Seríamos capaces de distinguir, desde el exterior, que no se trata del auténtico espacio-tiempo? Casi seguro que sí, porque en el espacio-tiempo en cuestión, los procesos físicos carecerían de continuidad. Los objetos cesarían de existir de repente en un momento dado, para reaparecer en otro. Y, lo que es aún más importante, las leyes de la física, o, cuando menos, las verdaderas leyes físicas, no serían respetadas. Deberían entrar en juego nuevas leyes que tuviesen en cuenta, explícita o implícitamente, que las instantáneas habían sido barajadas y que describiesen correctamente el nuevo espacio-tiempo.

Así pues, para nosotros la diferencia entre este nuevo espacio-tiempo y el real sería mayúscula. Pero ¿lo sería para sus habitantes? .potarían la diferencia? Nos estamos acercando aquí peligrosamente al absurdo, el familiar absurdo de la teoría de sentido común del tiempo. Pero siga conmigo sin impacientarse y sortearemos este peligro. *Por supuesto*, los habitantes del nuevo espacio-tiempo no podrían notar la diferencia. Si pudiesen, lo

demostrarían. Comentarían, por ejemplo, la existencia de discontinuidades en su mundo y publicarían trabajos científicos sobre ese tema (suponiendo, claro está, que pudiesen sobrevivir en semejante espacio-tiempo). Pero desde nuestro mágico observatorio podemos ver que sobreviven y publican trabajos científicos. No es posible, pues, consultarlos y comprobar que contienen únicamente observaciones sobre el espacio-tiempo original. Todos los registros de acontecimientos físicos dentro del espacio-tiempo, incluyendo los que se encuentran en las memorias y percepciones de los observadores conscientes, son idénticos a los del espacio-tiempo original. Tan sólo hemos barajado las instantáneas sin modificar su interior, de modo que sus habitantes siguen percibiendo el orden original.

Así pues, y en términos de la física real —la física percibida por los habitantes del espacio-tiempo—, tanto cortar, barajar y pegar el espacio-tiempo carece por completo de sentido. No tan sólo el nuevo espacio-tiempo, sino también toda la colección de instantáneas sin pegar, son físicamente idénticos al espacio-tiempo original. Representamos las instantáneas pegadas en el orden correcto porque así mostramos las relaciones entre ellas determinadas por las leyes de la física. Una imagen de las instantáneas pegadas en distinto orden seguiría representando los mismos acontecimientos físicos —la misma historia—, pero no mostraría correctamente las relaciones entre dichos acontecimientos. Las instantáneas tienen, pues, un orden *intrínseco*, definido por su contenido y por las leyes reales de la física. Cualquiera de las instantáneas, junto con las leyes de la física, determina no tan sólo lo que son todas las demás, sino el orden general y su propia posición en la secuencia. En otras palabras, cada instantánea lleva incorporado un «sello de tiempo», codificado en su contenido físico.

Así es como debe ser, si deseamos liberar al concepto del tiempo del error de invocar un marco global de tiempo ajeno a la realidad física. El sello de tiempo de cada instantánea es la lectura que da algún reloj natural existente en su universo. En algunas instantáneas —las que contienen la civilización humana, por ejemplo— hay auténticos relojes. En otras hay variables físicas —como la composición química del Sol, o de toda la materia del espacio— que pueden ser consideradas relojes, ya que toman valores claros y definidos en diferentes instantáneas de, al menos, una determinada región del espacio-tiempo. Podemos

estandarizarlas y calibrarlas para que se acoplen entre sí cuando se solapen.

Es posible reconstruir el espacio-tiempo utilizando el orden intrínseco determinado por las leyes de la física. Podemos empezar con cualquiera de las instantáneas. Calcularemos entonces cómo deberían ser las inmediatamente anterior y posterior, las buscaremos y, cuando las localicemos, las pegaremos a ambos lados de la instantánea original. La repetición de este proceso reconstruirá la totalidad del espacio-tiempo. Esos cálculos son demasiado complejos para ser realizados en la vida real, pero resultan válidos para un experimento mental durante el cual nos imaginamos desconectados del mundo físico real. (Además, estrictamente hablando, en física precuántica habría una infinita continuidad de instantáneas, de modo que el proceso que acabo de describir debería ser sustituido por un proceso limitador en el que el espacio-tiempo fuera ensamblado a lo largo de un número infinito de pasos; pero el principio seguiría siendo el mismo.)

La predecibilidad de un acontecimiento a partir de otro no implica que sean causa y efecto. Por ejemplo, la teoría de la electrodinámica dice que todos los electrones tienen una misma carga. Por consiguiente, utilizando dicha teoría podríamos —y a menudo lo hacemos— predecir el resultado de la medición de un electrón a partir el resultado de la medición de otro. Pero ninguno de ambos resultados ha sido *causado* por el otro. En realidad, por lo que sabemos, el valor de la carga de un electrón no ha sido causado por ningún proceso físico. Quizás sea «causado» por las mismas leyes de la física (si bien éstas, tal como las conocemos hoy día, no predicen la carga de un electrón, sólo dicen que todos los electrones tienen una misma carga). En cualquier caso, tenemos aquí un buen ejemplo de acontecimientos (resultados de mediciones de electrones) que son predecibles el uno a partir del otro, y viceversa, pero que no tienen efecto causal entre sí.

Veamos otro ejemplo. Si observamos dónde se encuentra una determinada pieza de un rompecabezas montado, conocemos las formas de todas las piezas restantes y sabemos que se hallan dispuestas del modo adecuado, podremos predecir dónde estarán. Pero ello no significa que la posición de las piezas restantes haya sido *causada* por la pieza cuya posición hemos observado. Dicha causalidad dependerá de cómo el rompecabezas llegó como un todo a su estado final. Si la pieza observada fue colocada al principio, estaba, sin duda, entre las causas de que las otras

piezas estuviesen donde están. Si fue otra pieza la que se colocó al principio, la situación de la pieza observada será un *efecto* de dicha colocación, no su causa. Pero si el rompecabezas fue creado por el golpe de un troquel que tenía las formas de todas las piezas y no ha sido desmontado desde entonces, ninguna de las posiciones de las piezas será causa o efecto de las demás. No fueron montadas en ningún orden, sino que fueron formadas simultáneamente en posiciones tales que obedecieran a las reglas del rompecabezas, reglas que convirtieron a esas posiciones en mutuamente predecibles. No obstante, ninguna de ellas causó las otras.

El deterninismo de las leyes físicas sobre los acontecimientos de un espacio-tiempo equivale a la predecibilidad de un rompecabezas que haya sido troquelado correctamente para que todas sus piezas encajen. Las leyes de la física determinan lo que sucede en un momento a partir de lo que ha sucedido en otro, del mismo modo que las reglas del rompecabezas determinan las posiciones de algunas piezas a partir de las de otras. Pero, al igual de lo que ocurre con el rompecabezas, que los acontecimientos sucedidos en diferentes momentos sean o no *causa* de otros dependerá de cómo llegaron los momentos a la posición que ocupan. No podemos decir, sólo con mirar el rompecabezas, que está allí porque fue montado pieza por pieza. Y, por lo que respecta al espacio-tiempo, sabemos que carece de sentido decir que un momento ha «sido montado» después de otro, ya que ello implicaría la existencia del transcurso del tiempo. Por lo tanto, sabemos que, si bien algunos acontecimientos pueden ser predecidos a partir de otros, ninguno causó a otro en el espacio-tiempo. Permítaseme subrayar de nuevo que ello es así según la física precuántica, en la que todo ocurre en el espacio-tiempo. Como vemos, el espacio-tiempo es incompatible con la existencia de causa y efecto. No es que la gente se equivoque al decir que determinados acontecimientos físicos son causas y efectos los unos de los otros, sino que dicha idea es incompatible con las leyes de la física del espacio-tiempo. Pero no tiene demasiada importancia, puesto que esa física es falsa.

Dije ya en el capítulo 8 que deben darse dos condiciones para que una entidad sea causa de su propia replicación: la primera es que la entidad sea efectivamente replicada, y la segunda es que la mayoría de sus variantes, en idénticas situaciones, no lo puedan ser. Esta definición da forma material a la idea de que una causa

es algo que provoca una diferencia en sus efectos, lo que resulta también de aplicación a la causalidad en general. Para que X sea la causa de Y, deben darse dos condiciones: la primera es que tanto X como Y sucedan, y la segunda es que Y no hubiese podido suceder de ser X distinta. Por ejemplo, la luz solar causó la vida sobre la Tierra porque ambas han sucedido en nuestro planeta *y* la vida no habría podido evolucionar en ausencia de la luz solar.

Así pues, el razonamiento sobre las causas y los efectos es igualmente inevitable para las variantes de dichas causas y efectos. Hablamos a menudo de lo que *habría* ocurrido si, permaneciendo otras cosas iguales, ciertos acontecimientos hubiesen sido distintos. Un historiador podría decir que, *«si* Faraday hubiese muerto en 1830, *entonces* la tecnología se habría retrasado unos veinte años más». El significado de esta afirmación parece perfectamente claro, y puesto que Faraday no murió en 1830, sino que descubrió la inducción electromagnética en 1831, parece también plausible. Equivale a decir que el desarrollo tecnológico posterior fue causado, en parte, por el descubrimiento de Faraday y, por consiguiente, también por su supervivencia. Pero ¿qué significa, en el contexto de la física del espacio-tiempo, razonar sobre el futuro de acontecimientos no existentes? Si no existe en el espacio-tiempo la muerte de Faraday en 1830, tampoco pueden existir sus consecuencias. Sin duda, podemos *imaginar* un espacio-tiempo que contenga dicho suceso, pero entonces, y puesto que tan sólo lo estamos imaginando, podemos imaginar también que contiene cualquier consecuencia que queramos. Podríamos imaginar, por ejemplo, que la muerte de Faraday provocó una *aceleración* del progreso tecnológico. Podríamos tratar de sortear esta ambigüedad imaginando tan sólo espacios-tiempo en los que, si bien el acontecimiento en cuestión es diferente del correspondiente en el espacio-tiempo real, las leyes físicas son las mismas. No hay ninguna razón evidente para restringir nuestra imaginación de este modo, pero, en cualquier caso, si las leyes de la física fueran las mismas, ello conllevaría que el suceso en cuestión *no hubiese podido* ser distinto, puesto que dichas leyes lo habrían determinado inequívocamente a partir de la historia previa. De modo que deberíamos imaginar que esa historia también era distinta. ¿Cuán distinta? El efecto de nuestra imaginaria variación de la historia depende de modo fundamental de lo que entendamos por «permaneciendo otras cosas iguales», y ello es de una terrible ambigüedad, puesto que son infinitos los

estados de cosas imaginables previas a 1830 que hubiesen podido conducir a la muerte de Faraday en dicho año. Algunos de ellos habrían provocado, indudablemente, un progreso tecnológico más rápido, y otros uno inás lento. ¿A cuáles de ellos nos referimos con el enunciado *«si... entonces ...»*? ¿Cuáles se acomodarían a la expresión «permaneciendo otras cosas iguales»? Por más que lo intentemos, no conseguiremos resolver esta ambigüedad dentro de la física del espacio-tiempo. No hay modo de evitar el hecho de que en el espacio-tiempo sólo una cosa sucede en la realidad, y todo lo demás es fantasía.

Nos vemos obligados a concluir que, en la física del espacio-tiempo, los enunciados condicionales cuya premisa es falsa («si Faraday hubiese muerto en 1830 ...») carecen de sentido. Los lógicos denominan a tales enunciados *condicionales contrafácticos*, *y* su status es el de una paradoja tradicional. Todos sabemos lo que significan tales enunciados; sin embargo, tan pronto como intentamos enunciar claramente su significado, éste parece evaporarse. El origen de esta paradoja no se encuentra ni en la lógica ni en la lingüística, sino en la física, en la falsa física del espacio-tiempo. La realidad física no es un espacio-tiempo, sino una entidad mucho mayor y más diversa: el *multiverso.* A primera vista, el multiverso se parece a un número muy grande de espacios-tiempo que coexistiesen e interactuasen ligeramente. Mientras que el espacio-tiempo es como una baraja de instantáneas, cada una de las cuales es la totalidad del espacio en un momento dado, el multiverso es como una vasta colección de tales barajas. Incluso esta (como veremos) ligeramente inexacta imagen del multiverso es susceptible ya de dar cabida a causas y efectos. Y es que en el multiverso hay, casi con toda seguridad, algunos universos en los que Faraday murió en 1830, y es un hecho (un hecho no observable, pero no por ello menos objetivo) si el progreso tecnológico se retrasó o no en ellos con respecto al nuestro. No hay nada arbitrario acerca de las variantes de nuestro universo a las que se refiere el condicional contrafáctico «si Faraday hubiese muerto en 1830 ...»: se refiere a *las variantes que ocurren realmente* en algún lugar del multiverso. La ambigüedad queda así resuelta. De nada sirve apelar a universos imaginarios, ya que podemos imaginar tantos como queramos y en las proporciones que deseemos. Pero en el multiverso los universos están presentes en proporciones definidas, de modo que tiene sentido afirmar que determinadas clases de sucesos son

«muy raras» o «muy comunes» en el multiverso, así como que algunos acontecimientos suceden a otros «en la mayoría de los casos». La mayoría de los universos lógicamente posibles no están presentes en el multiverso. Por ejemplo, no hay universos en los que la carga de un electrón difiera de la del nuestro, o en los que no sean de aplicación las leyes de la física cuántica. Las leyes de la física implícitamente aludidas en el condicional contrafáctico son las que están realmente en vigor en otros universos, a saber, las de la teoría cuántica. Por consiguiente, puede considerarse, sin la menor ambigüedad, que el enunciado *«si... entonces ...»*, significa «en la mayoría de los universos en los que Faraday murió en 1830, el progreso tecnológico se retrasó con respecto al nuestro». En general, podemos decir que un suceso X causa otro suceso Y en nuestro universo si tanto X como Y ocurren en él, pero en la mayor parte de las variantes de nuestro universo en las que X no ocurre, Y tampoco ocurre.

Si el multiverso fuese, literalmente, una colección de espacios-tiempo, el concepto cuántico del tiempo coincidiría con el clásico. Como muestra la figura 11.6, el tiempo seguiría siendo una secuencia de momentos. La única diferencia sería que, en un momento particular del multiverso, existirían múltiples universos en lugar de uno solo. En cada momento particular la realidad física sería entonces una «superinstantánea» consistente en instantáneas de múltiples versiones distintas de la totalidad del espacio. La totalidad de la realidad para la totalidad del tiempo está contenida en la serie de superinstantáneas, del mismo modo que en la física clásica lo estaba en la serie de instantáneas del espacio. A causa de la interferencia cuántica, cada instantánea no estaría ya determinada por completo por las precedentes en el mismo espacio-tiempo (si bien lo estaría bastante aproximadamente, ya que la física clásica constituye una buena aproximación de la física cuántica). Pero las superinstantáneas que empezasen con un momento particular quedarían completa y exactamente determinadas por las superinstantáneas precedentes. Este determinismo total no daría lugar a una absoluta predecibilidad, ni siquiera en principio, puesto que hacer una predicción requeriría conocer lo acaecido en todos los universos, y cada una de nuestras copias puede percibir directamente tan sólo un universo. No obstante, y por lo que concierne al concepto del tiempo, la imagen sería como la de un espacio-tiempo con una secuencia de momentos relacionados por leyes deterministas, sólo

que en cada momento ocurrirían más acontecimientos, aunque en su mayor parte estarían ocultos para cualquier copia de cualquier observador.

Sin embargo, el multiverso no es exactamente así. Desde hace décadas, obsesiona a la física teorética el deseo frustrado de enunciar una teoría cuántica del tiempo que sea operativa, la cual sería también la teoría cuántica de la gravedad. Sabemos ya, sin embargo, lo suficiente sobre ella para poder afirmar que, si bien las leyes de la física cuántica son perfectamente deterministas al nivel del multiverso, no lo dividen en espacios-tiempo separados, a la manera de la figura 11.6, ni tampoco en superinstantáneas, cada una de ellas determinante de las demás. Sabemos, pues, que el concepto clásico del tiempo como secuencia de momentos no puede ser cierto, si bien ofrece una buena aproximación en muchas circunstancias, es decir, en muchas regiones del multiverso.

Para esclarecer el concepto cuántico del tiempo, imaginemos que hemos cortado el multiverso en una serie de instantáneas individuales, como hicimos con el espacio-tiempo. ¿Con qué podremos volverlas a pegar? Igual que antes, las leyes de la física y las intrínsecas propiedades físicas de las propias instantáneas constituyen el único pegamento posible. Si el tiempo fuese en el multiverso una secuencia de momentos, debería ser posible identificar todas las instantáneas del espacio para un momento dado, a fin de agruparlas en una superinstantánea. De modo nada sorprendente, resulta que no hay manera de hacerlo. En el multiverso las instantáneas carecen de «sello de tiempo». No es posible preguntarse qué instantánea de otro universo sucede «en el mismo momento» como instantánea particular en el nuestro, puesto que ello implicaría, de nuevo, la existencia de un marco global de tiempo, exterior al multiverso, pero en relación al cual sucederían los acontecimientos en su interior. No existe tal marco.

No existe, por consiguiente, demarcación fundamental alguna entre instantáneas de otros tiempos e instantáneas de otros

universos. Este es el núcleo distintivo del concepto cuántico del tiempo: *Otros tiempos son simplemente casos especiales de otros universos.* Esta opinión surgió en los años sesenta, durante las primeras investigaciones acerca de la gravitación cuántica, en especial a partir de los trabajos de Bryce DeWitt, pero, por lo que sé, no fue formulada de modo general hasta 1983 por Don Page y William Wooters. Las instantáneas que denominamos «otros tiempos en nuestro universo» únicamente se distinguen de «otros universos» desde nuestra perspectiva, y tan sólo porque están especialmente relacionadas con las nuestras por las leyes de la física. Son, por lo tanto, aquellas de cuya existencia presenta mayor evidencia nuestra propia instantánea. Por esta razón las descubrimos miles de años antes de que descubriésemos el resto del multiverso, el cual, en comparación, nos afecta mucho más débilmente, mediante los efectos de interferencia. Elaboramos especiales estructuras gramaticales (formas verbales pasadas y futuras) para hablar de ellas. Elaboramos también otras estructuras (como enunciados de la clase *«si ... entonces ...»* y formas verbales condicionales y subjuntivas) para hablar de otras clases de instantáneas, sin saber siquiera si existían. Hemos situado tradicionalmente dos clases de instantánea —otros tiempos y otros universos— en categorías conceptuales completamente distintas. Vemos ahora que esa distinción resulta innecesaria.

Prosigamos ahora nuestra reconstrucción conceptual del multiverso. Tenemos a nuestra disposición muchas más instantáneas, un verdadero montón, pero empecemos, una vez más, con una instantánea individual de un universo en un momento dado. Si buscamos en ese montón otras instantáneas que sean muy parecidas a la original, veremos que se trata de instantáneas muy distintas de las que obtuvimos al cortar el espacio-tiempo, sobre todo, porque encontramos muchísimas que son absolutamente idénticas a la original. De hecho, cualquier instantánea está presente en infinidad de copias. No tiene, pues, sentido, preguntar cuántas instantáneas tienen, numéricamente, una determinada propiedad, sino *qué proporción* del total infinito la tiene. En aras de la brevedad, cuando hable de cierto «número» de universos, siempre me referiré a una determinada proporción del número total en el multiverso.

Si, además de las *variantes* de mí en otros universos, hay infinitas *copias* idénticas de mí, ¿cuál de ellas soy? Soy, por supuesto, todas ellas. Cada una de ellas acaba de formularse la

misma pregunta: «¿Cuál de ellas soy?», y cualquier manera cierta de contestarla deberá ofrecer la misma respuesta a todas ellas. Asumir que es físicamente significativo preguntarse «¿Cuál de las copias idénticas soy?», equivale a asumir que existe algún marco de referencia, fuera del multiverso, en relación al cual se podría contestar «El tercero por la izquierda...». Pero ¿qué «izquierda» podría ser ésa?, y ¿qué significa «el tercero»? Esta terminología sólo tiene sentido si imaginamos las instantáneas dispuestas en diferentes posiciones en algún espacio externo. Pero el multiverso no existe en ningún espacio externo ni en ningún tiempo externo: contiene todo el espacio y todo el tiempo existentes. Simplemente, existe, y es todo cuanto existe desde un punto de vista físico.

A diferencia de la física del espacio-tiempo, la teoría cuántica no determina, en general, lo que sucederá en una instantánea particular. En vez de ello, determina qué proporción de todas las instantáneas en el multiverso tendrá una determinada propiedad. Por esta razón, los que habitamos en el multiverso podemos, en ocasiones, realizar predicciones basadas únicamente en la posibilidad de nuestra propia experiencia, aunque lo que suceda en el multiverso esté completamente determinado. Supongamos, por ejemplo, que lanzamos una moneda al aire. Una predicción típica de la teoría cuántica sería que *si*, en un determinado número de instantáneas, ha sido lanzada al aire una moneda que gira de cierta manera cuando los relojes mostraban una determinada lectura, *entonces* habrá también un número de universos, equivalente a la mitad del número de instantáneas, en que los relojes mostrarán una lectura superior y la moneda habrá caído de «cara», y en un número de universos equivalente a la otra mitad del número de instantáneas los relojes también mostrarán una lectura superior y la moneda habrá caído de «cruz».

La figura 11.7 muestra la pequeña región del multiverso en la que acontecen esos sucesos. Incluso en una región tan pequeña hay una enorme cantidad de instantáneas que representar, por lo que tan sólo podemos destinar un punto del diagrama para cada una de ellas. Las instantáneas que observamos contienen relojes de una clase estándar, y el diagrama está dispuesto de manera que todas las instantáneas en que el reloj tiene una misma lectura aparezcan en una columna vertical, mientras que dichas lecturas se incrementan de izquierda a derecha. A medida que escaneamos a lo largo de cualquier línea vertical de puntos del diagrama, no todas las instantáneas que encontramos son distintas. Pasamos

por grupos de instantáneas idénticas, como indica el sombreado de la figura. Las instantáneas en que los relojes muestran la lectura más temprana se encuentran en el borde izquierdo del diagrama. Vemos que en todas esas instantáneas, que son idénticas, la moneda está en el aire. En el borde derecho del diagrama vemos que en la mitad de las instantáneas en las que los relojes muestran la lectura más tardía la moneda ha caído de «cara», mientras que en la otra mitad lo ha hecho de «cruz». En los universos con lecturas de reloj intermedias nos encontramos con tres tipos de universo, en proporciones que varían según la lectura de los relojes.

Si usted estuviera presente en la región de multiverso ilustrada, todas sus copias habrían visto girar la moneda en el aire al principio. Más tarde, la mitad de sus copias habrían visto caer la moneda de «cara», y la otra mitad, de «cruz». En algún estadio intermedio habría visto la moneda en un estado en el que se encontraba aún en movimiento, pero del que ya se podía predecir qué lado mostraría cuando quedara en reposo. La diferenciación de copias idénticas de un observador en versiones un tanto distintas es la responsable del carácter subjetivamente probabilista de las predicciones cuánticas. Si usted preguntase al principio qué resultado del lanzamiento de la moneda estaba destinado a presenciar, la respuesta sería que es estrictamente impredecible, puesto que la mitad de sus copias estarían destinadas a ver «cara» y la otra mitad a ver «cruz». No hay respuesta para la pregunta «qué mitad» verá una u otra cara de la moneda, al igual que no hay respuesta para la pregunta «¿Cuál soy?». A efectos prácticos, podríamos considerarlo como una predicción probabilista de que la moneda tiene un cincuenta por ciento de probabilidades de caer de «cara» y otro cincuenta por ciento de caer de «cruz».

El determinismo de la teoría cuántica, al igual que el de la física clásica, actúa hacia adelante y hacia atrás en el tiempo. El estado de la serie combinada de instantáneas con «caras» y

«cruces» del momento final de la figura 11.7 determina por completo el estado de «giro» del momento inicial, y viceversa. No obstante, y desde el punto de vista del observador, la información se pierde durante la trayectoria de la moneda, pues el observador puede experimentar el estado inicial, de «giro», de la moneda, pero es imposible que experimente su estado final de combinación de «caras» y «cruces». Por lo tanto, un observador podrá observar la moneda en estado inicial y predecir gracias a ello su estado futuro y las consiguientes probabilidades subjetivas, pero no es posible que ninguna de las copias posteriores de ese observador pueda observar la moneda y obtener la información necesaria para «retrodecir» el estado de «giro», ya que esa información estará para entonces distribuida por dos clases distintas de universo, lo que convierte en imposible toda «retrodicción» a partir del estado final de la moneda. Por ejemplo, si todo lo que sabemos es que la moneda muestra «cara», su estado inmediatamente anterior podría haber sido el que denomino de «giro», o podría haber estado girando en sentido contrario, o podría haber estado mostrando «cara» todo el tiempo. No hay aquí posibilidad alguna de «retrodicción», ni tan sólo probabilista. El estado inicial de la moneda, simplemente, no está determinado por el estado final de las instantáneas «cara», sino por el estado final conjunto de las instantáneas «cara» y «cruz».

Cualquier línea horizontal a través de la figura 11.7 atraviesa una secuencia de instantáneas con lecturas de reloj crecientes. Podríamos vernos tentados a considerar una línea así —por ejemplo, la que muestra la figura 11.8— como un espacio-tiempo y el diagrama como una serie de espacios-tiempo, uno por cada línea. Podemos leer en la figura lo que sucede en el «espacio-tiempo» definido por la línea horizontal. Durante un período, contiene a la moneda en giro. Luego, durante otro período, contiene a la moneda mientras se mueve de un modo que resultará predeciblemente en «cara». Pero, más adelante, contiene a la moneda en contradicción con el período anterior, mientras se mueve de un modo que resultará predeciblemente en «cruz», y el resultado acaba siendo, en efecto, «cruz». Pero se trata sólo de una deficiencia en el diagrama, como ya señalé en el capítulo 9 (véase figura 9.4, página 217). En un caso como el que nos ocupa, las leyes de la mecánica cuántica predicen que ningún observador que recuerde haber visto la moneda en el estado de «predeciblemente caras» puede verla en el de «predeciblemente

cruces»: ésta es la justificación fundamental para denominarlo estado de «predeciblemente caras». Por lo tanto, ningún observador en el multiverso reconocería los acontecimientos tal como suceden en el «espacio-tiempo» definido por la línea. Todo ello confirma que no podemos pegar las instantáneas de un modo arbitrario, sino únicamente de una manera que refleje sus interrelaciones determinadas por las leyes de la física. Las instantáneas a lo largo de la línea horizontal de la figura 11.8 no están lo suficientemente interrelacionadas para justificar su agrupación en un mismo universo. Sin duda, aparecen en el orden creciente de las lecturas del reloj, las cuales, en el *espacio-tiempo*, serían «sellos de tiempo» que bastarían para volver a montar el espacio-tiempo. Pero en el multiverso hay demasiadas instantáneas para que las lecturas del reloj basten por sí solas para situar una instantánea en relación con las demás. Para ello necesitamos tomar en consideración el intrincado detalle con que las instantáneas se determinan entre sí.

En la física del espacio-tiempo toda instantánea está determinada por cualquier otra. Como he dicho, en el multiverso esto no ocurre de modo general. Típicamente, el estado de un grupo de instantáneas idénticas (como aquellas en que la moneda está «girando») determina el estado de un segundo grupo con un número igual de instantáneas distintas entre sí (como las de «cara» y «cruz»). A causa de la propiedad de reversibilidad temporal de las leyes de la física cuántica, el estado más completo y con mayor número de valores de este último grupo determina también el estado del anterior. Sin embargo, hay regiones del multiverso y lugares del espacio en que las instantáneas de algunos objetos físicos quedan encadenadas durante cierto período y cada una de ellas determina a todas las demás hasta una buena aproximación.

Sucesivas instantáneas del sistema solar constituirían un buen ejemplo de ello. En esas regiones las leyes de la física clásica constituyen una buena aproximación a las de la física cuántica. En

tales regiones y lugares el multiverso se parece, en efecto, a la figura 11.6: es un conjunto de espacios-tiempo, y a ese nivel de aproximación el concepto cuántico del tiempo se reduce al clásico. Podemos distinguir, mediante aproximaciones, entre «tiempos distintos» y «universos distintos», y el tiempo es, por aproximación, una secuencia de momentos. Pero esa aproximación desaparece siempre al examinar más de cerca las instantáneas, o al mirar hacia adelante o hacia atrás en el tiempo, o a lo lejos en el multiverso.

Todos los resultados experimentales a los que tenemos acceso actualmente son compatibles con la aproximación de que el tiempo es una secuencia de momentos.

No es de esperar que dicha aproximación falle en ningún experimento terrestre previsible, pero la teoría nos dice que tiene que fallar, y estrepitosamente, en determinadas clases de procesos físicos, y, en primer lugar, en el primero de los cuales sería el propio Big Bang, el inicio del universo. Según la física clásica, el tiempo empezó en un momento en que el espacio era infinitamente denso y ocupaba tan sólo un único punto; antes de ese momento no había existido ningún otro. Según la teoría cuántica (hasta donde sabemos), las instantáneas muy próximas al Big Bang no están en ningún orden preciso. La propiedad secuencial del tiempo no empieza con el Big Bang, sino en algún momento posterior. Para lo que nos ocupa, carece de sentido preguntar cuánto después. Con todo, parece que los primeros momentos que pueden ser considerados, en una buena aproximación, como secuenciales ocurren, más o menos, cuando la física clásica extrapolaría que habían transcurrido 10^{-43} segundos (*el tiempo de Planck*) desde el Big Bang.

Se cree que se da un fallo similar de la secuencia del tiempo en el interior de los agujeros negros, y que se dará en el colapso final del universo (el Big Crunch), si es que ocurre. En ambos casos la materia es, o será, comprimida hasta una densidad infinita, según la física clásica, tal como se hallaba en el momento del Big Bang, y las fuerzas gravitatorias resultantes harán trizas la estructura del espacio-tiempo.

Por cierto, si le preocupa qué sucedió antes del Big Bang, o qué sucederá después del Big Crunch, ya tiene la respuesta. ¿Por qué cuesta tanto aceptar que no hubo momentos antes del Big Bang, ni los habrá después del Big Crunch, de modo que nada sucedió ni sucederá ni existió ni existirá? Porque es difícil imaginar

que el tiempo se detenga o se ponga en marcha. Pero, en realidad, no tiene que hacer ninguna de esas dos cosas, puesto que no se mueve. El multiverso no «empieza a existir» o «deja de existir»; estos términos presuponen el transcurso del tiempo. Es el hecho de que imaginamos que el tiempo transcurre lo que nos hace preguntarnos qué sucedió «antes» o sucederá «después» de la totalidad de la realidad.

También se considera que, a escala submicroscópica, los efectos cuánticos deforman y hacen trizas a su vez la estructura del espacio-tiempo, así como que a dicha escala existen bucles cerrados de tiempo (en realidad, minúsculas máquinas del tiempo). Como veremos en el próximo capítulo, esta clase de colapso de la secuencia del tiempo es también posible a gran escala, y está abierta la cuestión de si podría darse o no en las proximidades de objetos tales como los agujeros negros rotatorios.

Así pues, si bien no estamos aún en condiciones de detectar ninguno de esos efectos, nuestras teorías más fiables nos dicen ya que la física del espacio-tiempo no es nunca una descripción exacta de la realidad. Por muy buena aproximación que sea, el tiempo debe ser, en realidad, fundamentalmente distinto de la secuencia lineal que supone el sentido común. No obstante, todo en el multiverso está tan rígidamente determinado como en el espacio-tiempo clásico. Si extraemos una instantánea, las restantes la determinarán exactamente. Si extraemos la *mayoría*, las pocas que queden seguirán determinando a todas las extraídas, exactamente igual que en el espacio-tiempo. La única variación es que, a diferencia de este último, el multiverso no consiste en las series recíprocamente determinantes de instantáneas que he denominado superinstantáneas, las cuales podrían ser consideradas como los «momentos» del multiverso, sino que es un complejo rompecabezas multidimensional.

En este complejo rompecabezas que es el multiverso, que ni consiste en una secuencia de momentos ni permite el transcurso del tiempo, el concepto de causa y efecto de sentido común encaja a la perfección. El problema con el que nos encontrábamos en el espacio-tiempo, en relación con la causalidad, era que ésta constituye una propiedad tanto de las *variantes* de las causas y los efectos como de las propias causas y efectos. Puesto que estas variantes no existen en el espacio-tiempo, sino tan sólo en nuestra imaginación, nos enfrentamos a la incongruencia física de extraer conclusiones fundamentales de las propiedades imaginarias de

procesos físicos inexistentes («contrafácticas»). Pero en el multiverso existen las variantes, en distintas proporciones, y obedecen a leyes definidas y deterministas. Dadas estas leyes, constituye un hecho objetivo qué acontecimientos influyen en que ocurran otros y determinan cuáles serán. Supongamos que hay un grupo de instantáneas, no necesariamente idénticas, pero que comparten la propiedad X. Supongamos que, dada la existencia de ese grupo, las leyes de la física determinan que existe otro grupo de instantáneas que comparten la propiedad Y. Una de las condiciones para que X sea causa de Y ha sido satisfecha. La otra condición está relacionada con las variantes. Consideremos las variantes del primer grupo que carecen de la propiedad X. Si la existencia de éstas determina la de algunas de las instantáneas Y, X no fue la causa de Y, puesto que Y habría ocurrido aun sin X. Pero si tan sólo la existencia de las variantes no Y está determinada por el grupo de variantes no X, entonces X será causa de Y.

No hay nada en esta definición de causa y efecto que requiera lógicamente que las causas precedan a sus efectos, y podría suceder que no lo hicieran en situaciones muy anómalas, tales como cerca del Big Bang o en el interior de agujeros negros. En la experiencia cotidiana, sin embargo, las causas preceden siempre a sus efectos, lo cual se debe a que —al menos, en nuestra región del multiverso— el número de clases diferentes de instantánea tiende, por lo general, a incrementarse rápidamente con el tiempo, no a decrecer. Esta propiedad está relacionada con la segunda ley de la termodinámica, que enuncia que la energía ordenada —como la química o la potencial gravitatoria— puede ser transformada por completo en energía desordenada —como calor—, y viceversa. El calor es movimiento microscópico aleatorio. En términos del multiverso, esto significa múltiples estados de movimiento microscópicamente distintos en diferentes universos. Parece, por ejemplo, que, en sucesivas instantáneas de la moneda vistas en ampliaciones ordinarias, el proceso de caída transforma a un grupo de instantáneas «predeciblemente cara» idénticas en un grupo de instantáneas «cara» idénticas. Pero durante el proceso la energía del movimiento de la moneda se transforma en calor, de modo que en ampliaciones lo suficientemente grandes para poder apreciar moléculas individuales, las instantáneas del último grupo ya no son idénticas. Todas coinciden en que la moneda muestra la «cara», pero en cada una de ellas las moléculas de la moneda, del

aire que la rodea y de la superficie sobre la que descansa aparecen en múltiples configuraciones distintas. Por supuesto, las instantáneas iniciales «predeciblemente cara» tampoco son idénticas desde el punto de vista microscópico, puesto que también en ellas está presente algo de calor, pero la producción de éste durante el proceso implica que esas instantáneas sean mucho menos distintas entre sí que las del último grupo. Así, cada grupo homogéneo de instantáneas «predeciblemente cara» determina —y, por consiguiente, causa— la existencia de grandes números de instantáneas «cara» distintas desde el punto de vista microscópico. Pero ninguna instantánea «cara» determina por sí misma la existencia de instantáneas «predeciblemente cara», y, por consiguiente, no es su causa.

La transformación, en relación a cualquier observador, de posibilidades en realidades —de un futuro abierto a un pasado fijo—, encaja también a la perfección en este marco. Consideremos, una vez más, el ejemplo de la moneda lanzada al aire. Antes de lanzarla, el futuro está abierto desde el punto de vista de un observador, en el sentido de que aún es posible que cualquier resultado, «cara» o «cruz», sea observado por él. Desde su punto de vista, ambos resultados son posibilidades, aun cuando, objetivamente, ambos sean realidades. Cuando la moneda quede en reposo, las copias del observador se habrán diferenciado en dos grupos. Cada observador ha observado, y recuerda, tan sólo un resultado del lanzamiento de la moneda. De este modo, el resultado, una vez que pasa a formar parte del pasado del observador, se ha convertido en univalente y real para cada una de sus copias, aunque desde la perspectiva del multiverso sigue siendo tan ambivalente como siempre.

Permítaseme resumir los elementos del concepto cuántico del tiempo. El tiempo no es una secuencia de momentos ni transcurre. No obstante, nuestras ideas sobre sus propiedades son bastante correctas. Ciertos acontecimientos son, efectivamente, causas y efectos unos de otros. En relación con un observador, el futuro está, en efecto, abierto y el pasado fijo, y las posibilidades se transforman, en efecto, en realidades. La razón por la que nuestras teorías tradicionales sobre el tiempo son despropósitos es que tratan de explicar esas ideas correctas dentro del marco de una falsa física clásica. En la física cuántica esas teorías tienen sentido, puesto que el tiempo siempre ha sido un concepto cuántico. Existimos en múltiples versiones en universos

denominados «momentos». Cada versión de nosotros no es directamente consciente de las demás, pero tiene evidencia de su existencia porque las leyes físicas relacionan los contenidos de diferentes universos. Resulta tentador suponer que el momento del que somos conscientes es el único real, o al menos un poco más real que los demás, pero esto no es más que solipsismo. Todos los momentos son físicamente reales. La totalidad del multiverso es físicamente real. Nada más lo es.

TERMINOLOGÍA

Transcurso del tiempo. Supuesto desplazamiento del momento presente en dirección al futuro, o supuesto movimiento de nuestra conciencia de un momento a otro. Se trata de una idea carente de sentido.

Espacio-tiempo. El espacio y el tiempo considerados en conjunto como una entidad estática tetradimensional.

Física del espacio-tiempo. Teorías —como la de la relatividad— en las que se considera que la realidad es un espacio-tiempo. Puesto que la realidad es un multiverso, dichas teorías sólo pueden ser aproximaciones, en el mejor de los casos.

Libre albedrío. Capacidad de influir en los acontecimientos futuros de alguno de diversos modos posibles y escoger, además, cuál ocurrirá.

Condicional contrafáctico. Enunciado condicional cuya premisa es falsa (por ejemplo: «*Si* Faraday hubiese muerto en 1830, *entonces* habría ocurrido X»).

Instantánea, (únicamente en este capítulo). Un universo en un momento particular.

SUMARIO

El tiempo no transcurre. Otros tiempos son, simplemente,

casos especiales de otros universos.

Viajar en el tiempo podrá ser factible o no, pero disponemos ya de una concepción teórica razonablemente buena de cómo se desarrollaría el viaje en el tiempo en caso de serlo, concepción que involucra a las cuatro vías.

12. VIAJANDO EN EL TIEMPO

La idea de que el tiempo es, de algún modo, una cuarta dimensión adicional del espacio hace que resulte natural pensar que, si es posible desplazarse de un lugar a otro, quizás sea también posible viajar en el tiempo. Vimos en el capítulo anterior que la idea de «moverse» en el tiempo de la manera en que nos movemos en el espacio carece de sentido. No obstante, parece claro lo que queremos decir al hablar de viajar al siglo XXV o a la época de los dinosaurios. La ciencia ficción concibe a menudo las máquinas del tiempo como extravagantes vehículos en cuyos controles se introduce la fecha y la hora de la época que deseamos visitar (a veces se puede elegir incluso el lugar), se espera mientras el vehículo viaja hasta allí, y ya está. Si hemos escogido el lejano futuro, conversaremos con robots cuya inteligencia nos superará y nos maravillarán las naves interestelares, o bien (según las creencias políticas del autor) caminaremos entre ruinas radiactivas. Si, por el contrario, hemos escogido el lejano pasado, nos veremos obligados a defendernos de los ataques de un *Tyrannosaurus rex* mientras los pterodáctilos revolotean sobre nuestras cabezas.

La presencia de dinosaurios constituiría una impresionante evidencia de que realmente habíamos alcanzado una era anterior. Deberíamos comprobar esta evidencia y determinar la fecha con la mayor exactitud posible mediante la observación de algún «calendario» natural a largo plazo, como las configuraciones de las constelaciones en el cielo nocturno o las proporciones relativas de diversos elementos radiactivos en las rocas. La física nos proporciona muchos calendarios así, y sus leyes hacen que concuerden entre sí cuando se aplican adecuadamente. De acuerdo con la aproximación de que el multiverso es una serie de espacios-tiempo paralelos, cada uno consistente, a su vez, en una serie de «instantáneas» del espacio, la fecha definida de este modo es una propiedad de cada instantánea en su totalidad, *y* dos instantáneas cualesquiera están separadas por un intervalo temporal que es la diferencia entre sus fechas. Viajar en el tiempo es todo proceso que causa una disparidad entre, por una parte, dicho intervalo entre instantáneas y, por otra, nuestra propia experiencia del tiempo transcurrido mientras íbamos de la una a la otra. Podríamos referirnos a un reloj que transportáramos con

nosotros, o estimar cuánto tiempo hemos tenido ocasión de dedicar a pensar, o medir según criterios fisiológicos el envejecimiento de nuestros cuerpos. Si observamos que ha transcurrido mucho tiempo en el exterior, pero de acuerdo con todas las mediciones subjetivas hemos experimentado que el viaje ha sido más corto, ello indica que hemos viajado hacia el futuro. Si, en cambio, vemos que los relojes y los calendarios externos indican un tiempo particular y más tarde (subjetivamente) observamos que todos ellos, de manera constante, señalan un tiempo ante**rior,** habremos viajado hacia el pasado.

La mayor parte de los autores de ciencia ficción se dan cuenta de que viajar en el tiempo hacia el futuro es radicalmente distinto de hacerlo hacia el pasado. No prestaré aquí mucha atención al viaje en el tiempo hacia el futuro, ya que es, con mucho, la propuesta menos problemática. Incluso en la vida cotidiana, por ejemplo, cuando dormimos y nos despertamos, el tiempo subjetivamente experimentado puede ser más corto que el transcurrido en el exterior. Se podría decir que las personas que se recuperan de estados de coma que han durado varios años han viajado durante todo este tiempo hacia el futuro, si no fuera porque sus cuerpos habrán envejecido según el tiempo externo y no según el que han experimentado subjetivamente como transcurrido. Así, en principio, se podría emplear para todo el cuerpo una técnica similar a la que expuse en el capítulo 5 para ralentizar el cerebro de un usuario de realidad virtual y aplicarla al viaje en el tiempo hacia el futuro. La teoría especial de la relatividad de Einstein nos proporciona un método menos intrusivo al decir que, en general, un observador que acelera o decelera experimenta menos tiempo que **otro** que permanezca en reposo o en movimiento constante. Por ejemplo, un astronauta que realizara un viaje de ida y vuelta por el espacio que incluyera una aceleración a velocidades próximas a la de la luz experimentaría mucho menos tiempo que un observador que permaneciera en la Tierra. Este efecto se conoce con el nombre de *dilatación del tiempo.* Si le diéramos la suficiente aceleración, podríanlos hacer que la duración del vuelo fuera tan corta como quisiéramos desde el punto de vista del astronauta y tan larga como deseáramos según los parámetros de la Tierra. Así pues, podríamos viajar tan lejos en el futuro como gustáramos en un tiempo determinado, subjetivamente corto. Pero semejante viaje al futuro sería irreversible. El regreso requeriría viajar en dirección al pasado, y,

por mucho que se dilatara el tiempo, sería imposible que una nave espacial regresara de un vuelo antes de haberlo iniciado.

La realidad virtual y el viaje en el tiempo tienen, al menos, una cosa en común: ambos alteran sistemáticamente la relación habitual entre la realidad externa y la experiencia que de ésta tiene el usuario. Podríamos, pues, preguntarnos: si un generador universal de realidad virtual puede ser programado tan fácilmente para efectuar un viaje en el tiempo hacia el futuro, ¿existe algún modo de utilizarlo para viajar en el tiempo hacia el pasado? Si, por ejemplo, ralentizarnos nos proyecta hacia el futuro, ¿nos enviaría al pasado acelerarnos? La respuesta es *no.* Simplemente, nos parecería que el mundo exterior se relentizaba. Incluso en el caso límite inalcanzable en que el cerebro operase con rapidez infinita, el mundo exterior parecería haberse congelado en un momento particular. Eso seguiría siendo viajar en el tiempo, según la definición anterior, pero no hacia el pasado. Podríamos denominarlo viaje en el tiempo «hacia el presente». Recuerdo muy bien cuánto me hubiera gustado tener una máquina de viajar en el tiempo hacia el presente cuando tenía que estudiar para preparar un examen. ¿A qué estudiante no le habría gustado tenerla?

Antes de tratar el viaje en el tiempo hacia el pasado, ¿qué podemos decir sobre la *representación* de dicho viaje? ¿Hasta qué punto podría ser programado un generador de realidad virtual para proporcionar al usuario la experiencia de un viaje en el tiempo hacia el pasado? Veremos que la respuesta a esta pregunta, como todas las respuestas a preguntas que tratan del alcance de la realidad virtual, nos habla también de la realidad física.

Los aspectos distintivos de experimentar un entorno pasado son, por definición, experiencias de determinados objetos o procesos físicos —«relojes» y «calendarios»— en estados que ocurrieron únicamente en tiempos pasados (es decir, en instantáneas pasadas). Un generador de realidad virtual sería capaz, por supuesto, de representar dichos objetos en tales estados. Podría, por ejemplo, proporcionar la experiencia de vivir en la época de los dinosaurios, o en las trincheras de la Primera Guerra Mundial, y podría hacer que las constelaciones, las fechas de los periódicos, o lo que fuera, apareciesen de acuerdo con **esas épocas.** ¿De acuerdo hasta qué punto? ¿Existe algún límite para la exactitud con que podría ser representada una determinada época? El principio de Turing nos dice que un generador universal de realidad virtual puede ser construido y programado para

reproducir cualquier entorno físicamente posible, de modo que está claro que podría **ser** programado para representar cualquier entorno que haya existido físicamente antes.

Para representar una máquina del tiempo con un determinado repertorio de destinos pasados (y, por lo tanto, representar esos destinos), el programa debería incluir datos históricos de sus entornos. De hecho, necesitaría algo más que unos datos, ya que la experiencia del viaje en el tiempo conlleva mucho más que ver, simplemente, cómo **los** acontecimientos pasados se van desarrollando a nuestro alrededor. Reproducir ante el usuario grabaciones del pasado sería mera proyección de imágenes, no realidad virtual. Puesto que un auténtico viajero en el tiempo debe poder participar en los acontecimientos y actuar **sobre** un entorno del pasado, una representación fiel de una máquina del tiempo, como de cualquier otro entorno, debe ser interactiva. El programa debería calcular, para cada acción del usuario, la adecuada respuesta del entorno. Por ejemplo, para convencer al doctor Johnson de que una verdadera máquina del tiempo le había trasladado realmente a Roma, le debería resultar posible hacer algo más que observar de modo pasivo e invisible como pasaba por su lado Julio César. Debería poder comprobar la autenticidad de sus experiencias golpeando las piedras. Debería poder tocar a César, o, al menos, dirigirle la palabra en latín y obtener la correspondiente respuesta. Que una representación en realidad virtual de una máquina del tiempo sea fiel significa que debe responder a los procesos interactivos del mismo modo en que lo harían la máquina del tiempo real y los entornos a los que ésta viajara. Ello debería incluir, en nuestro caso, la representación de un Julio César que se comportara adecuadamente y hablara **en** latín.

Puesto que Julio César y la Roma antigua fueron objetos físicos, podrían, en principio, ser reproducidos con arbitraria fidelidad. No sería demasiado más difícil la representación de la pista central de Wimbledon, incluyendo a los espectadores. Por supuesto, la complejidad de los programas requeridos sería tremenda. Más compleja aún, o quizás incluso imposible, resultaría la tarea de recopilar la información necesaria para preparar los programas de representación de seres humanos específicos. Pero no pretendo tratar aquí de la programación ni me pregunto si podríamos averiguar lo suficiente sobre un entorno pasado (o, dado el caso, sobre un entorno presente o futuro) para diseñar el

programa que lo representara específicamente. Lo que me pregunto es si el *conjunto de todos los posibles* programas para generadores de realidad virtual incluye o no uno que ofrezca una representación en realidad virtual de un viaje en el tiempo hacia el pasado y, en caso afirmativo, cuán fiel es esa representación. Si *no* hubiera programas que representaran viajes en el tiempo, el principio de Turing implicaría que el viaje en el tiempo sería físicamente imposible, puesto que dicho principio nos dice que todo lo que es físicamente posible puede ser representado por algún programa. Al parecer, nos enfrentamos a un problema. Si bien hay programas capaces de representar fielmente entornos pasados, existen dificultades fundamentales para utilizarlos en la representación de viajes en el tiempo. Se trata de los mismos obstáculos que dificultan el propio viaje, es decir, las llamadas «paradojas» del viaje en el tiempo.

Veamos una de las más características de dichas paradojas. Construyo una máquina del tiempo y la utilizo para viajar hacia el pasado. Una vez allí, impido que mi yo anterior construya la máquina. Pero si la máquina no ha sido construida, no la podré haber usado para viajar hacia el pasado, por lo que no me habrá sido posible impedir su construcción. Así que ¿hago o no el viaje? Si lo hago, me privo de la máquina y, por consiguiente, no hago el viaje. Si no lo hago, me permito construir la máquina y emprender el viaje. Esta paradoja recibe en ocasiones la denominación de «paradoja del abuelo», y se formula en términos de utilización del viaje hacia el pasado para matar al propio abuelo antes de que tenga hijos. Si no tuvo hijos, menos pudo tener nietos; entonces, ¿quién lo mató? Estas dos variantes de la paradoja son las más comunes, y en ambas está presente un elemento de conflicto violento entre el viajero en el tiempo y gente del pasado, de modo que podríamos preguntarnos quién vencerá. Quizás el viajero no se salga con la suya y la paradoja sea evitada. Pero la violencia no es la parte esencial del problema. Si dispusiera de una máquina del tiempo, podría decidir lo siguiente: si hoy me visita mi futuro yo, procedente de mañana, mañana *no utilizaré* mi máquina del tiempo, mientras que si no recibo tal visita, mañana *utilizaré* la máquina para viajar de regreso a hoy y visitar a mi yo actual. Parece desprenderse de esta decisión que si utilizo la máquina, no la utilizaré, mientras que si no la utilizo, la utilizaré, lo cual es una contradicción.

Una contradicción indica una falsa asunción, de modo que

esas paradojas han sido consideradas tradicionalmente demostraciones de que el viaje en el tiempo es imposible. Otra asunción cuestionada a menudo es la del libre albedrío (si los viajeros en el tiempo pueden escoger o no entre diversos comportamientos, como es habitual en la vida real). Se llega a la conclusión de que si las máquinas del tiempo existiesen, el libre albedrío de los viajeros quedaría restringido. Podrían ser incapaces de tomar decisiones como la que he descrito, o, al viajar en el tiempo, tal vez olvidasen las decisiones tomadas antes de emprender el viaje. Pero en realidad, la falsa asunción que se esconde detrás de las paradojas no es ni la existencia de máquinas del tiempo ni la capacidad de los viajeros para comportarse o no del modo habitual. El defecto se encuentra en la teoría clásica del tiempo, que, como ya he demostrado, es, por diversas razones independientes de las que aquí nos ocupan, insostenible.

Si viajar en el tiempo fuera lógicamente imposible, lo sería también su representación mediante realidad virtual. Si requiriese una suspensión del libre albedrío, también la requeriría su representación mediante realidad virtual. Las paradojas del viaje en el tiempo pueden ser expresadas en términos de realidad virtual como sigue. La exactitud de una representación mediante realidad virtual es la fidelidad, hasta donde sea predecible, del entorno representado con el que se deseaba representar. En el caso del viaje en el tiempo, el entorno deseado ha existido históricamente. Pero tan pronto como el entorno representado responde, tal como se exige de él, a la acción del usuario, deja de ser históricamente exacto, ya que el entorno real nunca respondió al usuario, puesto que éste jamás influyó en él. Por ejemplo, el verdadero Julio César nunca se encontró con el doctor Johnson. Por consiguiente, el doctor Johnson, por el mero hecho de comprobar la fidelidad de la representación conversando con el César, destruiría esta fidelidad al crear un César históricamente inexacto. Una representación puede *comportarse* de modo correcto, si ofrece una imagen fiel de la historia, o *responder* con fidelidad a las acciones del usuario, pero no ambas cosas. Parecería, pues que, de un modo u otro, la representación mediante realidad virtual del viaje en el tiempo es intrínsecamente incapaz de ser exacta, lo que constituiría otra manera de decir que el viaje en el tiempo no puede ser representado mediante realidad virtual.

Pero ¿es este efecto, realmente, un impedimento para la

representación fiel del viaje en el tiempo? Por lo general, la imitación del comportamiento real de un entorno no es el objetivo de la realidad virtual; lo que de veras cuenta es que responda de modo adecuado. Tan pronto como empezamos a jugar al tenis sobre la representación de la pista central de Wimbledon, hacemos que su comportamiento difiera del de la auténtica pista, pero ello no hace que la representación sea menos fiel; antes al contrario, es lo requerido para que sea fiel. Fidelidad, en realidad virtual, significa el parecido del comportamiento representado con el que *mostraría* el entorno original de estar el usuario presente en él. Sólo al principio de la representación debe coincidir el estado del entorno representado con el del original. No es, por tanto, su estado, sino sus respuestas a las acciones del usuario, lo que debe ser fiel. ¿Por qué resulta esto paradójico para las representaciones de viajes en el tiempo y no para otras, como, por ejemplo, las de viajes corrientes?

Parece paradójico porque, en las representaciones de viajes en el tiempo hacia el pasado, el usuario tiene un papel singular, doble y, a veces, incluso múltiple. A causa de los recorridos implicados, en forma de espiral, que hacen que puedan coexistir e interactuar una o más copias del usuario, el generador de realidad virtual se ve, en efecto, requerido a *representar al usuario* al mismo tiempo que responde a sus actos. Imaginemos, por ejemplo, que soy el usuario de un generador de realidad virtual que ejecuta un programa de representación de viaje en el tiempo. Supongamos que, cuando pongo en marcha el programa, el entorno que veo a mi alrededor es un laboratorio futurista en cuyo centro se encuentra una puerta giratoria como las que vemos en los zaguanes de algunos edificios públicos, excepto que ésta es opaca y está casi por completo encerrada dentro de un cilindro igualmente opaco. El cilindro tiene un solo acceso, que sirve tanto de entrada como de salida. La puerta gira sin cesar en su interior. Parece, en principio, que no se puede hacer mucho más con este artilugio que entrar, dar una o más vueltas en la puerta giratoria, y salir. Pero sobre la entrada podemos leer: CAMINO AL PASADO. Es, pues, una máquina del tiempo ficticia, que representa el viaje en el tiempo mediante realidad virtual. Pero si existiese una auténtica máquina del tiempo para viajar al pasado, al igual que ésta, no sería una exótica clase de vehículo, sino de *lugar.* Más que ponernos a los mandos de un vehículo para conducirlo o pilotarlo hacia el pasado, seguiríamos cierta trayectoria por el

interior de ese lugar (quizás en una nave espacial convencional), para emerger en algún tiempo anterior.

En la pared del laboratorio simulado vemos un reloj que marca inicialmente las 12, y cerca de la entrada del cilindro hay un cartel con las instrucciones. Cuando termino de leerlas, son las 12.05, tanto según mi percepción como según el reloj. Las instrucciones dicen que si entro en el cilindro, doy una vuelta completa acompañando a la puerta giratoria y vuelvo a salir, en el laboratorio será cinco minutos más pronto. Me introduzco en uno de los compartimientos de la puerta giratoria. Mientras camino en círculo, mi compartimiento se cierra detrás de mí y unos momentos después alcanza de nuevo el hueco de acceso, por el que salgo. El laboratorio parece no haber cambiado, excepto por... ¿qué? ¿Qué es, exactamente, lo que debería esperar experimentar a continuación, si se trata de una reproducción fiel de un viaje en el tiempo hacia el pasado?

Retrocedamos un poco antes de contestar a esa pregunta. Supongamos que en la entrada hubiese un interruptor cuyas dos posiciones indicaran «interacción *activada*» e «interacción *desactivada*» y que, inicialmente, estuviera en la segunda. Ésta no permite al usuario participar en el pasado, sino tan sólo observarlo. En otras palabras, no proporciona una completa representación mediante realidad virtual del entorno pasado, sino tan sólo una generación de imagen.

Con esta disposición más sencilla, ya no hay ambigüedad o paradoja sobre qué imágenes deberían ser generadas cuando emerja de la puerta giratoria. Serán imágenes de mí en el laboratorio, haciendo lo que hacía a las 12. Una razón por la que no hay ambigüedad es que puedo recordar esos acontecimientos, de modo que puedo comparar las imágenes del pasado con mis propios recuerdos de lo ocurrido. Restringiendo nuestro análisis a un entorno pequeño y limitado a un breve período, evitaremos problemas tales como el de averiguar cómo era realmente Julio César, lo cual constituye un problema que tiene más que ver con

las posibilidades últimas de la arqueología que con las dificultades inherentes al viaje en el tiempo. En nuestro caso, el generador de realidad virtual puede obtener fácilmente la información que necesita para generar las imágenes requeridas: sólo tiene que grabar todo lo que hago. Es decir, no lo que hago en la realidad física (que no es otra cosa que permanecer tumbado dentro del generador de realidad virtual), sino en el entorno virtual del laboratorio. Así, en el momento en que emerjo de la máquina del tiempo, el generador de realidad virtual cesa de representar el laboratorio a las 12.05 y empieza a reproducir su grabación, que se inicia con las imágenes de lo que ocurrió a las 12. Me muestra la grabación con la perspectiva ajustada a mi posición actual y al lugar hacia donde miro, y reajusta dicha perspectiva del modo usual a medida que me muevo. Así pues, veo el reloj indicando de nuevo las 12, y mi yo anterior que lee el rótulo sobre la entrada y estudia las instrucciones, exactamente como hice cinco minutos antes (la situación está esquematizada en la figura 12.1). Veo a mi yo anterior, pero él no puede verme. No importa lo que haga, él —o, mejor dicho, *ello*, mi imagen en movimiento— no reacciona en modo alguno ante mi presencia. Transcurridos unos instantes, avanza hacia la máquina del tiempo.

Aunque me interponga y bloquee la entrada, mi imagen avanzará y entrará de todos modos, exactamente como hice yo, ya que si hiciera otra cosa sería una imagen inexacta. Hay varias maneras de programar un generador de realidad virtual para hacer frente a la situación en que la imagen de un objeto sólido debe pasar por la posición del usuario. Por ejemplo, la imagen podría atravesarlo como un fantasma, o podría empujar al usuario sin que éste pudiese resistirse. Esta última opción ofrecería una representación más real, ya que en ella las imágenes serían, de algún modo, táctiles además de visuales. No hay peligro de que mi imagen me dañe al apartarme, por más brusquedad con que lo haga, ya que, por supuesto, no estoy físicamente allí. En caso de no haber espacio suficiente para poderme apartar, el generador de realidad virtual podría hacer que me deslizase sin esfuerzo a través de una rendija, o incluso teleportarme más allá de algún obstáculo.

No es mi propia imagen lo único sobre lo cual ya no puedo influir. Puesto que hemos pasado temporalmente de la realidad virtual a la generación de imágenes, no podré modificar nada dentro del entorno simulado. Si hay un vaso de agua sobre una

mesa, ya no podré cogerlo y beber de él como antes de pasar por la puerta giratoria hacia el pasado simulado. Al solicitar una simulación de un viaje no interactivo en el tiempo hacia el pasado, lo cual es, en realidad, una reproducción de los acontecimientos específicos acaecidos cinco minutos antes, renuncio necesariamente al control sobre mi entorno. En realidad, cedo el control, por así decirlo, a mi yo anterior.

En cuanto mi imagen entra por la puerta giratoria, vuelven a ser, según el reloj, las 12.05, si bien han transcurrido diez minutos en la simulación, según mi percepción subjetiva. Lo que ocurra a partir de aquí, dependerá de lo que yo haga. Si, simplemente, permanezco de pie en el laboratorio, la próxima tarea del generador de realidad virtual deberá ser situarme en acontecimientos que ocurran después de las 12.05, hora del laboratorio. Carece aún de grabaciones de dichos acontecimientos, por la misma razón que yo no puedo recordarlos. En relación a mí, al laboratorio simulado y a la realidad física, tales acontecimientos no han sucedido aún, de modo que el generador de realidad virtual puede reanudar su representación completamente interactiva. El efecto neto será que habré estado durante cinco minutos en el pasado sin poder influir en él y luego habré regresado al «presente» que había abandonado, es decir, a la secuencia normal de acontecimientos en los que puedo influir.

También podría seguir a mi imagen al interior de la máquina del tiempo, viajar por su interior en su compañía y regresar de nuevo al laboratorio en el pasado. ¿Qué ocurriría? Una vez más, el reloj indica las 12, pero ahora puedo percibir *dos* imágenes de mi yo anterior. Una de ellas ve la máquina del tiempo por primera vez, y no puede vernos ni a mí ni a la otra imagen. La segunda imagen parece ver a la primera, pero no a mí, mientras que yo puedo verlas a las dos. Tan sólo la primera imagen parece influir en el entorno del laboratorio. En esta ocasión, y desde el punto de vista del generador de realidad virtual, no ha ocurrido nada especial en el momento del viaje en el tiempo. Se encuentra aún en el estado de «interacción *desactivada*», y, simplemente, continúa reproduciendo imágenes de acontecimientos ocurridos cinco minutos antes (desde mi punto de vista subjetivo), los cuales llegan ahora al momento en que empiezo a ver una imagen de mí.

Transcurridos otros cinco minutos, puedo volver a entrar, si lo deseo, en la máquina del tiempo, en esta ocasión acompañado por *dos* imágenes de mí (figura 12.2). Si repito el proceso,

aparecerá, tras cada cinco minutos subjetivos, una imagen de mí adicional. Cada una de ellas parecerá poder ver a las que la precedieron (según mi experiencia), pero no a las que la siguieron.

Si continúo la experiencia mientras me sea posible, el número máximo de copias de mí que puedan coexistir quedará limitado tan sólo por la estrategia de evitación de colisiones de que esté dotado el generador de imagen. Supongamos que éste trate de ponerme realmente difícil el entrar por la puerta giratoria en compañía de todas mis imágenes. Me veré entonces obligado a hacer algo distinto de viajar al pasado con ellas. Podría esperar un poco y tomar el compartimiento de la puerta giratoria siguiente al suyo, en cuyo caso llegaría al laboratorio un momento después que ellas. Pero eso sólo pospondría el problema de aglomeración en la máquina. Si me mantuviese girando en esa espiral, todos los pasillos aéreos para viajar en el tiempo dentro del período de cinco minutos después del mediodía estarían llenos, lo que me obligaría a llegar en un tiempo posterior, desde el cual no habría ya medio de retornar a dicho período. Esta es otra de las propiedades que las máquinas del tiempo deberían tener, en caso de existir físicamente. No ser sólo lugares, sino lugares con una capacidad limitada para permitir el tráfico hacia el pasado.

Otra consecuencia del hecho de que las máquinas del tiempo no sean vehículos, sino lugares o caminos, es que no tenemos absoluta libertad para escoger cuándo usarlas para viajar. Como demuestra el ejemplo anterior, sólo podemos utilizar una máquina del tiempo para viajar a tiempos y lugares en los que ha existido. Nunca podremos usarla para viajar a un tiempo anterior a su construcción.

El generador de realidad virtual dispone ahora de grabaciones de múltiples versiones distintas de lo que sucedió en el laboratorio entre las 12 y las 12.05. ¿Cuál de ellas describe la verdadera historia? No debería preocuparnos demasiado que no haya respuesta a esta pregunta, ya que plantea qué es real en una situación de la que hemos suprimido artificialmente la

interactividad, lo que hace que la prueba del doctor Johnson sea inaplicable. Podríamos argumentar que tan sólo la última versión, la que describe el máximo de copias de mí, es la verdadera, ya que todas las anteriores, en realidad, muestran la historia desde el punto de vista de personas que, debido a la regla artificial de no interacción, no podían ver lo que sucedía en su totalidad. Alternativamente, podríamos también defender que la primera versión de los acontecimientos, la que tiene una sola copia de mí, es la única real, puesto que es la única que he experimentado de manera interactiva. El aspecto fundamental de la no interactividad es que nos impedimos temporalmente cambiar el pasado, y puesto que todas las versiones subsiguientes difieren de la primera, no pueden describir el pasado. Sólo describen a alguien *que contempla* el pasado gracias a un generador universal de imágenes.

Se podría también argumentar que todas las versiones son reales por igual. Al fin y al cabo, cuando la experiencia haya terminado, no recordaré haber experimentado en el laboratorio una sola historia durante el período de cinco minutos, sino varias. Las experimenté sucesivamente, pero, desde la perspectiva del laboratorio, ocurrieron todas dentro del mismo período de cinco minutos. El registro total de mi experiencia requiere muchas instantáneas del laboratorio para cada momento definido por el reloj, en lugar de una instantánea por momento, que es lo habitual. En otras palabras, se trata de una representación de universos paralelos. Es esta última interpretación la que resulta más próxima a la verdad, como podremos comprobar repitiendo el experimento, pero esta vez con la interacción activada.

Lo primero que deseo manifestar acerca del modo interactivo, en el que tengo libertad para influir en el entorno, es que *una* de las cosas que puedo decidir que sucedan es, exactamente, la secuencia de acontecimientos que acabo de describir para el modo no interactivo. Es decir, puedo retroceder, encontrarme con una o más copias de mí y comportarme (si soy lo bastante buen actor) como si no pudiera verlas. Sin embargo, debo observarlas con cuidado. Si deseo recrear la secuencia de acontecimientos que ocurrió cuando realicé el experimento con la interacción desactivada, debo recordar lo que hicieron mis copias, a fin de poderlo repetir en mis visitas subsiguientes a ese tiempo.

Al principio de la sesión, cuando veo por primera vez la máquina del tiempo, aparecen inmediatamente una o más copias

de mí. ¿Por qué? Porque, con la interacción activada, cuando utilice la máquina a las 12.05 podré influir en el pasado al que regreso, y este pasado es el que sucede ahora, a las 12. Así pues, mis futuros yo o yoes llegan para ejercitar su derecho a influir en el laboratorio a las 12, influirme y, en particular, ser vistos por mí.

Las copias de mí se dedican a sus asuntos. Consideremos la tarea calculatoria que deberá realizar el generador de realidad virtual para representar esas copias. Hay ahora un elemento adicional que la hace insuperablemente más difícil de lo que era en el modo no interactivo. ¿Cómo podrá el generador de realidad virtual averiguar lo que van a hacer las copias de mí? Carece aún de grabaciones de dicha información, ya que, en tiempo físico, la sesión acaba de empezar. Sin embargo, debe ofrecerme inmediatamente representaciones de mi futuro yo.

Mientras siga decidido a simular que no las puedo ver y a imitar después lo que hacen, no van a estar sometidas a pruebas de fidelidad demasiado estrictas. El generador de realidad virtual sólo deberá obligarlas a hacer *algo* que yo pudiese hacer o, más exactamente, a comportarse de cualquier modo que yo pudiese imitar. Dada la tecnología en la que suponemos que se basa el generador, podemos presumir que ello no debería superar sus posibilidades. Dispone de un modelo matemático exacto de mi cuerpo y tiene, en cierta medida, acceso directo a mi cerebro. Podría aprovecharlo para calcular algún comportamiento que yo pueda imitar y hacer que sus primeras representaciones de mí obren de ese modo.

Empiezo, pues, la experiencia viendo cómo por la puerta giratoria salen algunas copias de mí y hacen algo. Simulo no verlas y, cinco minutos después, entro y salgo por la puerta giratoria e imito lo que he visto que hacía antes la primera copia. Cinco minutos más tarde vuelvo a entrar y salir e imito a la segunda copia, y así sucesivamente. Mientras tanto, me doy cuenta de que una de las copias repite siempre lo que *yo* he estado haciendo durante los primeros cinco minutos. Al final de la secuencia de viajes en el tiempo, el generador de realidad virtual volverá a disponer de varios registros de lo que ocurrió durante los cinco minutos después de mediodía, pero esta vez todas las grabaciones serán idénticas. En otras palabras, sólo ha ocurrido una historia, a saber, que me encontré con mi futuro yo pero pretendí no darme cuenta. Más tarde, me convertí en ese futuro yo y viajé al pasado para encontrarme con mi yo anterior, para el cual, al parecer, pasé

inadvertido. Todo ello resulta muy exacto y no tiene nada de paradójico... pero es irreal. Se consiguió mediante un intrincado juego de referencias mutuas entre el generador de realidad virtual y yo: lo imitaba mientras él me imitaba. Pero con las interacciones normales activadas ya no tendré opción a jugar de ese modo.

Si yo tuviera realmente acceso al viaje en el tiempo en realidad virtual, desearía, sin duda, comprobar la autenticidad de la representación. En el caso que analizamos, la comprobación empezaría tan pronto como viera mis copias. Lejos de ignorarlas, entablaría inmediatamente conversación con ellas. Estoy en condiciones mucho mejores para verificar su autenticidad de lo que lo estaría el doctor Johnson para comprobar la de Julio César. Para poder superar esa prueba inicial, las reproducciones de mí deberán ser efectivamente seres inteligentes artificiales, más aún, seres tan parecidos a mí, al menos en sus respuestas a estímulos externos, que me convenzan de que son reproducciones exactas de mí tal como podría ser dentro de cinco minutos. El generador de realidad virtual debe ejecutar programas similares a mi mente en contenido y complejidad. Una vez más, no es la dificultad de diseñar tales programas el tema que nos ocupa. Estamos investigando el principio del viaje en el tiempo en realidad virtual, no su factibilidad. No nos incumbe dónde consiga sus programas nuestro hipotético generador de realidad virtual. Lo que preguntamos es si el *conjunto de todos los posibles* programas incluye o no uno que represente fielmente el viaje en el tiempo. Nuestro generador de realidad virtual dispone, en principio, de los medios necesarios para averiguar todos mis posibles comportamientos en diversas situaciones. Esta información se encuentra localizada en el estado físico de mi cerebro, y, en principio, mediciones lo bastante precisas podrían leerla. Un método (probablemente inaceptable) para conseguirlo sería que el generador de realidad virtual pudiera hacer que mi cerebro interactuara en realidad virtual con un entorno de prueba, grabar su comportamiento y luego volver al estado inicial, quizás «rebobinando» el programa. La razón de que este método sea probablemente inaceptable estriba en que cabe presumir que yo *experimentaría* el entorno de prueba, y, aunque no debería recordarlo después, exigiría del generador de realidad virtual que me proporcionase las experiencias que especificase y no otras.

En cualquier caso, lo que interesa para nuestros propósitos es que, puesto que mi cerebro es un objeto físico, el principio de

Turing nos dice que se encuentra dentro del repertorio de un generador universal de realidad virtual. Así pues, debería resultar posible que la copia de mí superara la prueba de la autenticidad de su parecido conmigo. Pero ésta no es la única prueba que deseo realizar. Lo que me interesa, sobre todo, es comprobar si el viaje en el tiempo en sí es representado auténticamente. A tal fin, deseo averiguar no tan sólo si esa persona es en realidad yo, sino también si viene de veras del futuro. Puedo comprobarlo, en parte, interrogándola. Debería responderme que recuerda haber estado en mi lugar hace cinco minutos, y que luego entró y salió por la puerta giratoria para encontrarse conmigo. Yo debería advertir también que esa persona trata de comprobar *mi* autenticidad. ¿Por qué debería esa copia de mí hacer tal cosa? Porque el modo más estricto y directo por el que puedo comprobar su parecido con mi yo futuro sería esperar hasta que yo haya pasado por la máquina del tiempo, y averiguar entonces dos cosas: en primer lugar, si la copia de mí con la que me encuentro se comporta como yo recuerdo haberme comportado, y en segundo lugar, si *yo* me comporto como recuerdo que se comportó *mi copia.*

Sin duda, en ambos casos la representación no superaría la prueba de la autenticidad. Sólo con proponérmelo, conseguiré comportarme de modo diferente a como se comportó mi copia. Hacer que ella se comporte, a su vez, de modo diferente a como yo me comporté resultaría igual de fácil: todo lo que debería hacer sería formularle una pregunta que a mí, en su lugar, no se me hubiera formulado y que tuviese una respuesta característica. De este modo, por mucho que se me parezcan en aspecto y personalidad, los seres que salen de la máquina del tiempo de realidad virtual no son auténticas representaciones de la persona que seré en breve. En realidad, nunca lo podrían ser, puesto que tengo la firme intención de no comportarme como ellas cuando sea mi turno de utilizar la máquina, y, ya que el generador de realidad virtual me permite ahora interactuar libremente con el entorno representado, nada me impedirá cumplir dicha intención.

Recapitulemos. Cuando empieza el experimento, me encuentro con una persona que, en todos los aspectos, soy yo, pequeñas variaciones aparte. Esas variaciones indican de modo evidente que viene del futuro: recuerda el laboratorio a las 12.05, tiempo que, desde mi perspectiva, todavía no ha sucedido. Recuerda haber echado a andar a esa hora, haber pasado por la puerta giratoria y haber llegado a las 12. Y, antes de que ocurriera

todo eso, recuerda haber iniciado el experimento a las 12 y haber visto por primera vez la puerta giratoria, de la que salían copias de sí mismo. Afirma que esto sucedió hace más de cinco minutos según su percepción subjetiva, si bien, de acuerdo con la mía, todo el experimento ha durado menos de cinco minutos. Y así sucesivamente. Sin embargo, aunque supera todas las pruebas de ser una versión de mí en el futuro, puedo demostrar que no es *yo* en el futuro. Cuando compruebo si es la persona específica en la que me convertiré en el futuro, no supera la prueba. De modo similar, me dice que *yo* no supero la prueba de ser él en el pasado porque no hago exactamente lo que recuerda haber hecho.

Así, cuando viajo al pasado del laboratorio, encuentro que no es el mismo pasado del que acabo de llegar. Debido a su interacción conmigo, la copia de mí con la que me encuentro allí no se comporta exactamente como recuerdo haberme comportado. Por lo tanto, si el generador de realidad virtual tuviera que registrar todo lo ocurrido durante esa secuencia de viaje en el tiempo, debería almacenar de nuevo varias instantáneas para cada momento definido según el reloj del laboratorio, pero en esta ocasión serían todas distintas. En otras palabras, existirían varias historias distintas paralelas del laboratorio durante el período de cinco minutos del viaje en el tiempo. Una vez más, he experimentado cada una de dichas historias por separado, pero en esta ocasión las he experimentado todas interactivamente, de modo que no hay excusa para decir que ninguna de ellas sea menos real que las demás. Así pues, lo que está siendo representado es un pequeño multiverso. Si se tratara de un viaje físico en el tiempo, las múltiples instantáneas para cada momento serían universos paralelos. Dado el concepto cuántico del tiempo, eso no debería sorprendernos. Sabemos que las instantáneas que forman, aproximadamente, una sola secuencia del tiempo en nuestra experiencia cotidiana son, en realidad, universos paralelos. No experimentamos, por lo general, los otros universos paralelos que existen al *mismo* tiempo, pero tenemos buenas razones para pensar que son reales. Así que, si encontrásemos un modo, aún por especificar, de viajar a un tiempo anterior, ¿por qué deberíamos esperar que dicho método transportara por fuerza a cada copia de nosotros precisamente a la misma instantánea que dicha copia experimentó en su momento? ¿Por qué deberíamos esperar que cada visitante que recibamos del futuro nos salude desde la particular instantánea futura en la que

nos encontraremos cuando llegue el momento? No hay razón para esperarlo. Pedir interactuar con el entorno pasado significa pedir el poder de modificarlo, lo que significa, por definición, pedir encontrarse en una instantánea pasada distinta de la que recordamos. Un viajero en el tiempo sólo podría regresar a la misma instantánea (o, lo que quizás sea lo mismo, a una instantánea idéntica) en el caso, extremadamente limitado, que he analizado antes, en que no existe ninguna interacción efectiva ente las copias que se encuentran y en el que el viajero consigue que todas las historias paralelas sean idénticas.

Vamos a someter ahora a la máquina del tiempo en realidad virtual a la última comprobación. Me prepararé para escenificar una paradoja. Decido tomar la resolución que insinué antes que podía adoptar: decido que, si sale de la máquina del tiempo una copia de mí a las 12, *no entraré* en ella a las 12.05 ni a ninguna otra hora durante el experimento. Pero, si no sale ninguna de mí, *entraré* en ella a las 12.05, saldré a las 12 y no la volveré a usar. ¿Qué ocurrirá? ¿Saldrá alguien de la máquina o no? Sí... ¡y no! Depende del universo acerca del cual estemos hablando. Recordemos que ocurre más de una cosa en el laboratorio a las 12. Supongamos que no veo salir a nadie de la máquina, como ilustra el punto señalado como «Inicio» a la derecha de la figura 12.3. Luego, y perseverando en mi resolución, espero hasta las 12.05 y entro y salgo por la ya familiar puerta giratoria. Al salir de ella a las 12 me encuentro, como es natural, con otra versión de mí, de pie en el punto señalado como «Inicio» a la *izquierda* de la figura 12.3. Mientras conversamos, descubrimos que ambos tenemos la misma resolución. Por consiguiente, y ya que he salido dentro de su universo, se comportará de modo distinto a como lo hice yo. Obrar consecuentemente con su intención le lleva a *no* utilizar la máquina del tiempo. A partir de entonces, él y yo podemos continuar interactuando en tanto que dure la simulación, y habrá dos versiones de mí en ese universo. En el universo del que procedo el laboratorio permanecerá vacío a partir de las 12.05, ya que nunca volveré a él. No hemos escenificado ninguna paradoja. Ambas versiones de mí han conseguido realizar nuestra resolución compartida, que, después de todo, no ha resultado lógicamente imposible de llevar a cabo.

En este experimento, mi álter ego y yo hemos tenido experiencias distintas. Él vio salir a alguien de la máquina del tiempo a las 12 y yo no. Nuestras experiencias hubieran sido igualmente fieles a nuestra resolución, e idénticamente no paradójicas, de haber intercambiado nuestros papeles. Es decir, yo podría haberle visto salir de la máquina a las 12 y no haberla utilizado después. En ese caso, ambos habríamos acabado en el universo en que yo empecé. En el universo en que él empezó, el laboratorio permanecería vacío.

¿Cuál de estas dos posibilidades autoconsistentes me mostrará el generador de realidad virtual? Durante esta representación de un proceso intrínsecamente multiversal, yo interpreto sólo a una de mis copias, mientras que el programa se ocupa de la otra. Al principio del experimento, las dos copias parecen idénticas (si bien en la realidad física son distintas, ya que tan sólo una de ellas está conectada a un cuerpo y un cerebro físicos, fuera del entorno virtual). Pero en la versión física del experimento —caso de existir una máquina del tiempo físicamente—, los dos universos que contuviesen las copias de mí destinadas a encontrarse serían al principio absolutamente idénticos y ambas copias igualmente reales. En el momento del multiverso en que nos encontráramos (en un universo) o no (en el otro), ambas copias quedarían diferenciadas. Carece de sentido preguntar *qué* copia de mí tendría una determinada experiencia: mientras seamos idénticos, no es posible diferenciarnos. Los universos paralelos no disponen de números de serie secretos, se distinguen tan sólo por lo que sucede en ellos. Por consiguiente, al realizar su representación para una de mis copias, el generador de realidad virtual debe recrear para mí el efecto de existir en dos copias idénticas, que se diferencian después y tienen experiencias distintas. Puede causar ese efecto, literalmente, escogiendo de modo aleatorio, con igual número de probabilidades, cuál de los dos papeles interpretará (y, por lo tanto, y según mi resolución, cuál interpretaré yo). Escoger aleatoriamente equivale a lanzar al aire una versión electrónica de una moneda imparcial, es decir, que muestre «cara» en la mitad de los universos en que sea lanzada y «cruz» en la otra mitad. Así pues, en la mitad de los

universos interpretaré un papel, y en la otra mitad, el otro. Esto es, exactamente, lo que ocurriría con una máquina del tiempo real.

Hemos visto que la capacidad de un generador de realidad virtual para representar fielmente el viaje en el tiempo depende de que disponga de información detallada sobre el estado mental del usuario. Esto podría inducirnos a preguntarnos si las paradojas han sido realmente evitadas. Si el generador de realidad virtual conoce de antemano lo que voy a hacer, ¿soy de veras libre para realizar las comprobaciones que desee? No hace falta que profundicemos aquí en la naturaleza del libre albedrío. Soy, sin duda, libre para hacer lo que me plazca en este experimento, en el sentido de que sea cual sea la manera en la que decida reaccionar ante el pasado simulado —incluso aleatoriamente, si éste es mi deseo—, el generador de realidad virtual me lo permite y todos los entornos con los que interactúo se ven afectados por mis actos y sus reacciones influyen sobre mí, exactamente del mismo modo en que lo harían si no estuviese teniendo lugar el viaje en el tiempo.

La razón por la que el generador de realidad virtual necesita información procedente de mi cerebro no es predecir *mis* actos, sino representar el comportamiento de mis copias en otros universos. Su problema consiste en que, en la versión real de esa situación, habría copias de mí en universos paralelos, inicialmente idénticas y que, por lo tanto, tendrían las mismas propensiones y tomarían las mismas decisiones que yo. (Más lejos en el multiverso, habría otras versiones de mí que serían ya distintas al final del experimento, pero una máquina del tiempo nunca podría causar mi encuentro con ellas.) Si hubiera algún otro modo de representar a estas personas, el generador de realidad virtual no necesitaría información alguna procedente de mi cerebro, ni tampoco los prodigiosos recursos calculatorios de que hemos hablado. Si, por ejemplo, algunas personas que me conocieran bien pudiesen imitarme hasta un cierto grado de exactitud (dejando aparte algunos atributos externos tales como la apariencia y el tono de voz, que son relativamente fáciles de reproducir), el generador de realidad virtual podría utilizarlas para desempeñar los papeles de mis copias en universos paralelos y representar, en consecuencia, el viaje en el tiempo con ese nivel de exactitud.

Una máquina del tiempo real no se enfrentaría, por supuesto, a estos problemas. Proporcionaría, simplemente, caminos por los que yo y mis copias, que ya existían, nos pudiésemos encontrar, y

no constreñiría nuestros comportamientos, ni nuestras interacciones al encontrarnos. Las maneras en que los caminos se interconectan —es decir, a qué instantáneas nos conduciría la máquina del tiempo— se verían afectadas por mi estado físico, incluyendo mi estado mental.

Ello no difiere de la situación habitual, en la que mi estado físico, reflejado en mi propensión a comportarme de diversas maneras, influye en lo que ocurre. La gran diferencia entre la experiencia de la que estamos hablando y la experiencia cotidiana estriba en que cada una de mis copias tiene un gran efecto potencial sobre otros universos (al viajar a ellos).

¿Equivale realmente a viajar en el tiempo el poder viajar al pasado de otros universos, pero no al del nuestro? ¿Tendría, quizás, más sentido hablar de viaje *interuniversal* que de viaje en el tiempo? No. Los procesos que he descrito son realmente viajar en el tiempo. En primer lugar, no es cierto que *no podamos* viajar a una instantánea en la que ya hemos estado. Si disponemos las cosas de manera adecuada, nos es posible. Por supuesto, si cambiamos algo del pasado —si lo hacemos distinto de como era cuando estuvimos en él—, nos encontraremos en un pasado distinto. Un verdadero viaje en el tiempo nos permitiría cambiar el pasado. En otras palabras, nos permitiría hacer el pasado distinto de cómo lo recordamos (en este universo), lo cual significa distinto de cómo es realmente en las instantáneas en las que no llegamos a cambiar nada, las cuales incluyen, por definición, aquellas en que recordamos haber estado.

Así pues, desear cambiar las instantáneas pasadas específicas en las que estuvimos alguna vez carece por completo de sentido. Pero esa falta de sentido no tiene nada que ver con el viaje en el tiempo. Es un despropósito que procede directamente de la absurda teoría clásica del transcurso del tiempo. Cambiar el pasado significa escoger en qué instantánea pasada estar, no transformar una instantánea específica en otra. En este sentido, cambiar el pasado no difiere de cambiar el futuro, que es algo que hacemos constantemente. Cada vez que hacemos una elección, cambiamos el futuro: lo hacemos distinto de como hubiera sido en caso de haber escogido otra opción. Semejante idea no tendría sentido en la física clásica del espacio-tiempo, con su futuro único determinado por el presente, pero sí lo tiene en la física cuántica. Cuando elegimos, cambiamos el futuro, que ya no será lo que hubiese sido en los universos en que escogemos de otro modo.

Pero en ningún caso cambia una instantánea determinada del futuro. Y es que no podría cambiar, puesto que no hay ningún transcurso del tiempo con respecto al cual pudiera cambiar. «Cambiar» el futuro significa escoger en qué instantánea estaremos, y «cambiar» el pasado significa, exactamente, lo mismo. Puesto que no hay transcurso del tiempo, no es posible cambiar ninguna instantánea del pasado, por ejemplo, alguna en la que recordemos haber estado. No obstante, si de algún modo tuviésemos acceso físico al pasado, no hay razón por la que no pudiésemos cambiarlo exactamente del mismo modo en que cambiamos al futuro, es decir, escogiendo estar en una instantánea distinta de aquella en que hubiésemos estado de haber escogido de otro modo.

Razonar acerca de la realidad virtual ayuda a la comprensión del viaje en el tiempo, ya que el concepto de realidad virtual requiere que nos tomemos en serio los «acontecimientos contrafácticos», y, en consecuencia, el concepto multiversal cuántico del tiempo parece natural cuando es representado mediante realidad virtual. Al ver que el viaje en el tiempo hacia el pasado forma parte del repertorio de un generador universal de realidad virtual, nos damos cuenta de que la idea de dicho viaje es perfectamente sensata. Pero ello no significa afirmar que sea posible hacerla realidad en el plano físico. Después de todo, el viaje a velocidad superior a la de la luz, las máquinas de movimiento perpetuo y muchas otras cosas imposibles según las leyes de la física pueden representarse mediante realidad virtual. Ningún razonamiento sobre la realidad virtual podrá demostrar nunca que determinado proceso está permitido por las leyes de la física (si bien podría demostrar que no lo está: si hubiésemos llegado a la conclusión contraria, ello habría implicado, de acuerdo con el principio de Turing, que el viaje en el tiempo no podría ocurrir físicamente). Así pues, ¿qué nos dicen acerca de la física nuestras conclusiones positivas sobre el viaje en el tiempo en realidad virtual?

Nos dicen cómo sería el viaje en el tiempo en caso de ocurrir. Nos dicen que el viaje en el tiempo hacia el pasado sería, inevitablemente, un proceso que afectaría a varios universos, interconectados e interactuantes. En dicho proceso los participantes viajarían, en general, entre universos siempre que viajaran en el tiempo. Los modos precisos en que los universos estuviesen conectados dependerían, entre otras cosas, de los

estados mentales de los participantes.

Así pues, para que el viaje en el tiempo sea posible, es indispensable la existencia de un multiverso. Es igualmente necesario que las leyes de la física que gobiernen dicho multiverso sean tales que, en presencia de una máquina del tiempo y de potenciales viajeros en él, los universos queden interconectados en el modo que he descrito y no en otro. Por ejemplo, si decido no usar la máquina, no debe poder aparecer en mi instantánea ninguna versión de mí viajando en el tiempo; es decir, ningún universo en que las versiones de mí utilicen la máquina del dempo debe poder conectar con el mío. Y en el caso de que decida usarla, mi universo deberá conectar con otro en el que también la use. Y si voy a intentar escenificar una «paradoja», como hemos visto, mi universo deberá establecer contacto con otro en el que una copia de mí tenga la misma intención que yo, pero, al llevarla a cabo, acabe comportándose de modo distinto a mí. Hay que subrayar que todo esto coincide con lo que predice la teoría cuántica. En pocas palabras, el resultado es que, si existen caminos hacia el pasado, los viajeros que transiten por ellos deben ser libres de interactuar con el entorno del mismo modo en que lo podrían hacer si dichos caminos no condujesen al pasado. En ningún caso resulta inconsistente el viaje en el tiempo ni se impone especiales limitaciones al comportamiento de los viajeros.

Esto nos lleva a la pregunta de si es posible la existencia de caminos al pasado. Esta cuestión ha sido objeto de mucha investigación y sigue estando abierta. El punto de partida habitual es una serie de ecuaciones que constituyen la base (predictiva) de la teoría gereral de la relatividad de Einstein, actualmente la mejor teoría del espacio y el tiempo de que disponemos. Estas ecuaciones, conocidas como *ecuaciones de Einstein*, tienen múltiples soluciones, cada una de las cuales describe una posible configuración tetradimensional de espacio, tiempo y gravedad. Las ecuaciones de Einstein permiten, ciertamente, la existencia de caminos al pasado, y han sido encontradas para ellas muchas soluciones con dicha propiedad. Hasta hace poco, la práctica habitual ha sido ignorar sistemáticamente esas soluciones, pero ello no era consecuencia de ninguna razón emergente de la propia teoría, ni de ningún razonamiento físico. Era consecuencia de que los físicos tenían el convencimiento de que el viaje en el tiempo «conduciría a paradojas» y, por consiguiente, las ecuaciones de Einstein debían ser «no físicas». Lo arbitrario de esta segunda

suposición trae a la memoria lo ocurrido en los primeros años que siguieron a la introducción de la teoría de la relatividad general, cuando las soluciones que describían el Big Bang y un universo en expansión fueron rechazadas por el propio Einstein. Trató de cambiar esas ecuaciones para que describiesen un universo estático. Más tarde se referiría a ello como el error más grande de su vida. La expansión del universo fue comprobada experimentalmente por el astrónomo estadounidense Edwin Hubble. De modo parecido, las soluciones a las ecuaciones de Einstein obtenidas por el astrónomo alemán Karl Schwarzschild, que fueron las primeras en describir agujeros negros, fueron erróneamente rechazadas por «no físicas». Describían un fenómeno contraintuitivo: una región de la que, en principio, es imposible escapar y unas fuerzas gravitatorias que resultan infinitas en el centro del agujero negro. La idea dominante en la actualidad es que los agujeros negros existen efectivamente y tienen las propiedades predichas por las ecuaciones de Einstein.

Tomadas al pie de la letra, las ecuaciones de Einstein predicen que el viaje en el tiempo sería posible en las proximidades de objetos masivos en rotación, tales como los agujeros negros, a condición de que éstos giren lo bastante deprisa y se den determinadas condiciones complementarias. Muchos físicos, sin embargo, dudan de que estas predicciones sean realistas. No se conocen agujeros negros que giren tan deprisa como sería necesario, y se ha argumentado (de modo poco convincente) que no parece posible acelerar alguno de ellos artificialmente, puesto que cualquier materia en rápida rotación que se lanzara hacia él sería despedida y no podría penetrar en su interior. Quizás los escépticos tengan razón, pero mientras su resistencia a aceptar la posibilidad del viaje en el tiempo se base en la creencia de que conduce a paradojas, no estará justificada.

Aun en el caso de que se llegase a comprender mejor las ecuaciones de Einstein, no parece que pudiesen proporcionar respuestas concluyentes a la cuestión del viaje en el tiempo. La teoría general de la relatividad es anterior a la teoría cuántica, y no resulta completamente compatible con ella. Nadie ha conseguido aún formular una versión cuántica satisfactoria, es decir, una teoría cuántica de la gravedad. Sin embargo, y según los argumentos que he expuesto, los efectos cuánticos serían dominantes en situaciones de viaje en el tiempo. Las propuestas candidatas a teoría cuántica de la gravedad se caracterizan porque

no sólo permiten la existencia de conexiones dirigidas al pasado en el multiverso, sino que predicen que dichas conexiones se forman y se rompen continuamente de modo espontáneo. Esto sucede tanto en el espacio como en el tiempo, pero sólo a escala submicroscópica. Según algunas hipótesis, el camino típico formado por estos efectos tendría una longitud de 10^{-35} metros, permanecería practicable durante un tiempo de Planck (10^{-43} segundos), y por lo tanto, sólo se adentraría, aproximadamente, un tiempo de Planck en el pasado.

El viaje en el tiempo hacia el futuro, que, esencialmente, requeriría tan sólo cohetes eficientes, se encuentra en un horizonte tecnológico más bien distante, pero que parece alcanzable. El viaje en el tiempo hacia el pasado, en cambio, requiere la manipulación de agujeros negros o alguna otra violenta disrupción gravitatoria de la estructura del espacio y el tiempo, y, si llega a ser practicable, lo será en un futuro muy remoto. Por el momento, no conocemos nada en las leyes de la física que descarte el viaje en el tiempo hacia el pasado, es más, lo hacen verosímil. Los descubrimientos futuros en física fundamental podrían cambiar esta situación. Podría descubrirse que las fluctuaciones cuánticas en espacio y tiempo se vuelven insuperablemente fuertes cerca de las máquinas del tiempo y sellan de modo irremediable sus entradas (Stephen Hawking, por ejemplo, ha expuesto que algunos de sus cálculos sugieren esa posibilidad, pero eso está por confirmar). O quizás algún fenómeno aún desconocido descarte el viaje en el tiempo hacia el pasado, o proporcione algún método nuevo y más fácil para hacerlo factible. No podemos predecir el futuro crecimiento del conocimiento, pero si el desarrollo de la física fundamental continúa permitiendo el viaje en el tiempo en principio, su realización práctica se convertirá, con toda seguridad, en un mero problema tecnológico que se resolverá en un momento u otro.

Puesto que ninguna máquina del tiempo proporciona acceso a tiempos anteriores al momento en que inició su existencia, y a causa de que la teoría cuántica dice que están interconectados los universos, hay ciertos límites para lo que podemos esperar aprender mediante el uso de dichas máquinas. Hasta que hayamos construido una no podemos esperar la llegada de visitantes o, al menos, de mensajes procedentes del futuro y que salgan de ella. ¿Qué nos dirán? Algo que, con toda seguridad, no nos traerán son noticias de nuestro propio futuro. La pesadilla determinista de la

profecía de una futura catástrofe inescapable, acontecida a pesar de —o quizás a causa de— nuestros esfuerzos por evitarla sólo es tema para mitos y ciencia ficción. Los visitantes del futuro no pueden conocer nuestro futuro mejor que nosotros, ya que no proceden de él. Sin embargo, nos pueden hablar del futuro de su universo, cuyo pasado fue idéntico al nuestro. Pueden traer consigo noticiarios y programas de actualidad grabados, así como periódicos atrasados con fechas que empiezan a partir de mañana. Si su sociedad tomó alguna decisión equivocada que la condujo al desastre, nos podrán advertir. Nosotros podremos seguir o no su consejo. Si lo seguimos, tal vez evitemos caer en el mismo error o bien —no hay garantía posible— descubramos que el resultado de nuestra elección es aún más catastrófico que el de la suya.

En conjunto, no obstante, deberíamos podernos beneficiar bastante del estudio de su futura historia. Si bien no es la nuestra, y a pesar de que conocer la inminencia de algún desastre no equivale a saber qué hacer para evitarlo, es presumible que podríamos aprender mucho de lo que, desde nuestro punto de vista, *sería posible* que ocurriera.

Nuestros visitantes podrían darnos a conocer grandes logros científicos y artísticos. Si éstos hubieran sucedido en el futuro próximo del otro universo, sería probable que existieran en el nuestro copias de las personas que los realizaron trabajando para conseguirlos. De golpe, se encontrarían ante versiones acabadas de sus trabajos. ¿Estarían agradecidos? He aquí otra aparente paradoja del viaje en el tiempo. Puesto que no parece crear inconsistencias, sino curiosidades, ha sido más tratada en la literatura de ficción que en los debates científicos como argumentación en contra del viaje en el tiempo (si bien algunos filósofos, como Michael Dummett, la han tomado en serio). La denomino *paradoja del conocimiento* del viaje en el tiempo y se desarrolla como sigue. Un historiador futuro interesado en Shakespeare utiliza una máquina del tiempo para visitar al gran autor mientras que está escribiendo *Hamlet.* Mantienen una conversación, en el transcurso de la cual el viajero en el tiempo muestra a Shakespeare el texto del soliloquio «Ser o no ser...» de Hamlet, que ha traído consigo desde el futuro. A Shakespeare le gusta, y lo incorpora a la obra. En otra versión, Shakespeare muere y el viajero asume su identidad y alcanza la gloria fingiendo que escribe sus obras, cuando en realidad lo único que hace es copiarlas en secreto de un ejemplar de las obras completas de

Shakespeare que se trajo del futuro. En otra versión, el viajero en el tiempo se sorprende de no poder encontrar a Shakespeare. Por una serie de casualidades, tiene que adoptar la personalidad de Shakespeare y, de nuevo, plagiar sus obras. Le coge gusto a esa vida y, años después, se da cuenta de que *él* es Shakespeare: nunca ha existido otro.

Incidentalmente, la máquina del tiempo de estas historias debería haber sido suministrada por alguna civilización extraterrestre que hubiera desarrollado ya el viaje en el tiempo en la época de Shakespeare, y que estuviese dispuesta a prestar a nuestro historiador uno de sus escasos y no renovables pasillos aéreos para viajar en el pasado a aquella época. O quizás (lo cual me parece todavía menos probable) hubiera alguna máquina del tiempo disponible que se encontrara de modo natural en las cercanías de algún agujero negro.

Todas estas historias relatan una cadena —o, más bien, un círculo— de acontecimientos perfectamente consistente. La razón de que resulten chocantes y merezcan la denominación de paradojas ha de buscarse en el hecho de que en cada una de ellas la gran literatura nace sin que nadie la haya escrito ni creado, y este hecho, si bien consistente desde un punto de vista lógico, contradice profundamente nuestra comprensión de los orígenes del conocimiento. Según los principios epistemológicos que expuse en el capítulo 3, el conocimiento no surge nunca completamente formado. Existe sólo como resultado de procesos creativos que son, paso a paso, procesos evolutivos, que empiezan con un problema y avanzan mediante las nuevas teorías provisionales, las críticas y la eliminación de errores hacia un problema-situación nuevo y preferible. Es así como escribió Shakespeare sus obras. Es así como descubrió Einstein sus ecuaciones de campo. Es así como cada uno de nosotros logra resolver cualquier problema, pequeño o grande, en su vida o crea obras artísticas de valor.

Y así es también como se originan nuevas especies vivientes. El «problema» es, en este caso, un nicho ecológico. Las «teorías» son los genes, y las nuevas teorías provisionales, los genes mutados. La «crítica» y la «eliminación de errores» están representadas ahora por la selección natural. El conocimiento es generado mediante la acción humana intencionada, pero las adaptaciones biológicas lo son por un ciego mecanismo inconsciente. Las palabras que empleamos para describir ambos procesos son distintas, como distintos son también físicamente,

pero ambos obedecen detalladamente a las mismas leyes de la epistemología. En un caso se denominan teoría de Popper del desarrollo del conocimiento científico, y en el otro, teoría de Darwin de la evolución. También es posible formular una paradoja del conocimiento en términos de especies vivas. Supongamos que transportamos en una máquina del tiempo a algunos mamíferos a la era de los dinosaurios, anterior a la evolución de aquéllos. Soltamos a nuestros animales, los dinosaurios se extinguen y los mamíferos ocupan su lugar. De este modo, se han originado nuevas especies sin haber evolucionado. Resulta aún más fácil ver por qué esta versión es filosóficamente inaceptable: implica un origen no darwiniano de las especies y, de manera específica, el *creacionismo.* Es cierto que no se invoca aquí a ningún creador, en el sentido tradicional de la palabra. Sin embargo, el origen de las especies en esta historia es claramente sobrenatural: no ofrece ninguna explicación *—y descarta la posibilidad de que exista alguna explicación—* sobre cómo ocurrieron las complejas y específicas adaptaciones de las especies a sus nichos.

Así pues, las situaciones de la clase paradoja del conocimiento infringen los principios epistemológicos o, si se quiere, evolutivos. Son paradójicas tan sólo porque conllevan la creación, desde la nada, de complejos conocimientos humanos y complejas adaptaciones biológicas. Historias análogas que involucren en la espiral a otras clases de objetos o informaciones no son paradójicas. Observemos un guijarro en la playa, viajemos después a ayer, busquemos el guijarro hasta encontrarlo en otro lugar y llevémoslo adonde lo hemos encontrado hoy. ¿Por qué lo hemos encontrado allí? Porque lo hemos transportado a ese lugar. ¿Por qué lo hemos transportado allí? Porque lo hemos encontrado en ese lugar. Hemos causado que una información (la situación del guijarro) cobre vida en una espiral autoconsistente. ¿Y qué? El guijarro tenía que estar en alguna parte. Mientras la historia no implique hacer algo de la nada, bien sea el conocimiento o la adaptación, no será una paradoja.

En la perspectiva del multiverso, el viajero en el tiempo que visita a Shakespeare no viene del futuro de esa copia de Shakespeare. Puede influir en la copia a la que visita, e incluso reemplazarla. Pero nunca podrá visitar a la copia que existió en el universo del que él ha partido. Y *ésa* es la copia que escribió Sus obras. Así pues, dichas obras tuvieron un auténtico autor. Las espirales paradójicas de la clase planteada en la historia no

existen. El conocimiento y la adaptación —incluso en la presencia de caminos al pasado— no surgen de repente, sino que aumentan gradualmente gracias a la creatividad humana o a la evolución biológica, y no puede ser de otro modo.

Desearía poder afirmar que esta exigencia es también rigurosamente aplicada por las leyes que la teoría cuántica impone al multiverso. Espero que así sea, pero resulta difícil de demostrar, puesto que esa tan deseada propiedad no es fácil de expresar mediante el lenguaje corriente de la física teórica. ¿Qué fórmula matemática distingue el «conocimiento» o la «adaptación» de la información carente de valor? ¿Qué atributos físicos distinguen a un proceso «creativo» de otro que no lo sea? Si bien no estamos aún en condiciones de responder a estas preguntas, no creo que la situación sea desesperada. Recordemos las conclusiones del capítulo 8 acerca de la trascendencia de la vida y el conocimiento en el multiverso. Señalé allí (por razones que nada tienen que ver con el viaje en el tiempo) que la creación de conocimiento y la evolución biológica son procesos significativos. Una de las razones para ello era que dichos procesos —y sólo ellos— tienen un efecto particular sobre los universos paralelos, es decir, la creación *es* una estructura transuniversal que los hace iguales. Algún día, cuando comprendamos los detalles de este efecto, estaremos en condiciones de definir el conocimiento, la adaptación, la creatividad y la evolución en términos de convergencia de universos.

Cuando «escenifico una paradoja», hay eventualmente dos copias de mí en un universo y ninguna en el otro. Es una regla general que, después de que haya tenido lugar el viaje en el tiempo, el número total de copias de mí, contado en todos los universos, no deberá haber cambiado. De modo semejante, las leyes corrientes de conservación de la masa, la energía y otras magnitudes físicas continúan siendo de aplicación para el multiverso en conjunto, si bien no necesariamente para todos los universos particulares. Sin embargo, no hay una ley de conservación del conocimiento. La posesión de una máquina del tiempo nos proporcionaría acceso a un conocimiento de procedencia completamente distinta: la creatividad de las mentes de otros universos. Éstas podrían, a su vez, recibir conocimiento de nosotros, por lo que podríamos perfectamente hablar de un «intercambio» de conocimiento —y por supuesto, de artefactos que incorporen conocimiento— a través de múltiples universos.

Pero no hay que tomarse esta analogía al pie de la letra. El multiverso nunca será una zona de libre intercambio, ya que las leyes de la física cuántica imponen serias restricciones a las interconexiones entre instantáneas. En primer lugar, los universos se conectan inicialmente tan sólo en el momento en que son idénticos, pero el hecho mismo de estar conectados hace que empiecen a diferenciarse. Sólo cuando esas diferencias se hayan acumulado, y haya sido creado nuevo conocimiento en un universo y enviado viajando hacia el pasado al otro, podremos recibir conocimiento que no exista aún en el nuestro.

Una manera más acertada de ver este «intercambio» de conocimiento sería considerar todos nuestros procesos generadores de conocimiento, toda nuestra cultura y civilización, todos los procesos mentales de todos los individuos, e incluso toda la biosfera en evolución, como un gigantesco *cálculo* en el que todo ese conjunto ejecutaría un programa informático automotivado y autogenerado. Más específicamente, se trataría, como ya he mencionado, de un programa de realidad virtual en el acto de representar, con creciente fidelidad, la totalidad de la existencia. En otros universos hay otras versiones de ese generador de realidad virtual, algunas idénticas, otras muy diferentes. Si dicho generador de realidad virtual tuviera acceso a una máquina del tiempo, podría recibir algunos de los resultados de los cálculos realizados por sus contrapartidas en otros universos, en la medida en que las leyes de la física permitiesen el necesario intercambio de información. Cada fragmento de conocimiento obtenido de la máquina del tiempo tendría un autor específico en algún lugar del multiverso, pero beneficiaría a un número indecible de distintos universos. Una máquina del tiempo es, pues, un recurso calculatorio que permite que ciertas clases de cálculos puedan ser realizadas con una eficiencia mucho mayor de lo que lo serían por un ordenador individual. Alcanza esta eficiencia compartiendo las tareas calculatorias con sus propias copias en universos diferentes.

En ausencia de máquinas del tiempo, tiende a haber muy poco intercambio de información entre universos, porque las leyes de la física predicen en este caso muy poco contacto causal entre ellos. Puede afirmarse, con un buen nivel de aproximación, que el conocimiento creado en una serie de instantáneas idénticas alcanza a relativamente pocas otras instantáneas, a saber, las que están agrupadas en espacios-tiempo situados en el futuro de las

instantáneas originales. Pero eso es tan sólo una aproximación. Los fenómenos de interferencia son el resultado del contacto causal entre universos próximos. Ya vimos en el capítulo 9 que incluso ese minúsculo nivel de contacto puede ser utilizado para intercambiar información significativa y calculatoriamente útil entre universos.

El estudio del viaje en el tiempo nos proporciona un campo —si bien, de momento, sólo teorético y al nivel de experimento mental en el que podemos leer en grandes caracteres algunas de las conexiones entre las que denomino las «cuatro grandes vías». Todas ellas desempeñan papeles esenciales en la explicación del viaje en el tiempo. Éste podrá ser factible algún día o no, pero, si lo es, no exigirá ningún cambio fundamental en la concepción del mundo de nadie, o, al menos, no de la de quienes compartan ampliamente la que expongo en este libro. Todas las conexiones que el viaje en el tiempo podría establecer entre pasado y futuro son comprensibles y no paradójicas. Y todas las conexiones que necesitaría entre campos aparentemente desconectados del conocimiento existen ya.

TERMINOLOGÍA

Viaje en el tiempo. Tan sólo puede considerarse como tal el dirigido hacia el pasado. En éste el viajero experimenta el mismo instante, definido según relojes y calendarios externos, varias veces en sucesión subjetiva, mientras que en el viaje en el tiempo hacia el futuro el viajero llega a un instante posterior en un tiempo subjetivamente más corto que el definido según relojes y calendarios externos.

Máquina del tiempo. Objeto físico que capacita al usuario para viajar al pasado. Es más adecuado considerarla un lugar, o un camino, que un vehículo.

Paradoja del viaje en el tiempo. Situación aparentemente imposible que un viajero en el tiempo podría hacer realidad si fuera posible viajar en él.

Paradoja del abuelo. Situación en que la persona que viaja al pasado, una vez en él, realiza acciones cuyas consecuencias le

impedirán efectuar ese viaje (es decir, altera el curso de la historia).

Paradoja del conocimiento. Aquella en que **se** crea conocimiento de la nada mediante el viaje en el tiempo.

SUMARIO

No sabemos si algún día será posible viajar en el tiempo, pero no es algo paradójico. Al viajar al pasado conservamos nuestra libertad habitual de acción, pero, en general, acabamos en el pasado de un universo diferente. El estudio del viaje en el tiempo es una disciplina teorética en la que las cuatro vías principales tienen significación por igual: la mecánica cuántica, gracias a sus universos paralelos y su concepto cuántico del tiempo; la teoría de la calculabilidad, a causa de las conexiones entre realidad virtual y viaje en el tiempo, y porque las características distintivas del viaje en el tiempo pueden ser analizadas como nuevas modalidades de cálculo; en cuanto a la epistemología y la teoría de la evolución, su significación estriba en las limitaciones que imponen en los modos de adquirir el conocimiento.

No sólo están relacionadas las cuatro vías como parte de la estructura de la realidad, sino que se dan también notables paralelismos entre los cuatro campos de conocimiento como tales. Las cuatro teorías básicas gozan del poco habitual status de ser a un mismo tiempo aceptadas y rechazadas, utilizadas y no creídas por la mayor por la mayor parte de quienes trabajan en dichos campos.

13. LAS CUATRO VÍAS

Un estereotipo ampliamente extendido del proceso científico es el del joven innovador e idealista que se enfrenta a los viejos carcamales del *establishment* científico. Los vejestorios, instalados en la confortable ortodoxia de la que se han convertido al mismo tiempo en defensores y prisioneros, se escandalizan ante cualquier desafío que la cuestione. Rehúsan escuchar la voz de la crítica, enzarzarse en discusiones o aceptar la evidencia, y tratan de anular las ideas del innovador.

Este estereotipo ha sido convertido en filosofía por Thomas Kuhn, autor del influyente libro *The Structure of Scientific Revolutions.* Según él, el *establishment* científico está definido por la creencia de sus miembros en una serie de teorías dominantes que, en conjunto, forman una determinada concepción del mundo o *paradigma.* Un paradigma es un aparato psicológico y teórico mediante el cual los miembros del *establishment* observan y explican todo lo concerniente a su experiencia. (Dentro de cada campo específico de conocimiento, la física, por ejemplo, podemos hablar también de un «paradigma» específico para él.) En caso de surgir alguna observación que pareciera infringir el paradigma en vigor, los miembros del *establishment*, simplemente, harían oídos sordos. Cuando se ven forzados a enfrentarse a la evidencia de que la teoría que sustenta ese paradigma ha sido infringida, tienden a considerarlo una «anomalía», un error experimental, un fraude o cualquier otra cosa que les permita mantener su paradigma incólume. Así pues, Kuhn cree que los valores científicos de apertura a la crítica y a la provisionalidad en la aceptación de teorías, así como los métodos científicos de comprobación experimental y de abandono de teorías dominantes cuando quedan refutadas, son, en esencia, mitos, humanamente imposibles de concretar cuando está en juego algún tema científico fundamental.

Kuhn acepta que, en el caso de asuntos científicos *insignificantes*, sucede algo parecido al proceso científico (tal como lo he descrito en el capítulo 3). Opina que la ciencia avanza mediante alternancia de eras: existen eras de «ciencia normal» y eras de «ciencia revolucionaria». Durante una era de ciencia normal, casi todos los científicos comulgan con las teorías fundamentales dominantes y se esfuerzan para que sus

observaciones y teorías subsidiarias encajen en el paradigma en vigor. Su investigación consiste básicamente en atar cabos sueltos, mejorar las aplicaciones prácticas de las teorías, clasificar, reformular *y* confirmar. Cuando resulta posible, pueden incluso aplicar métodos científicos en el sentido popperiano de la palabra, pero nunca descubren nada fundamental, porque no cuestionan nada fundamental. Llegan luego algunos jóvenes alborotadores que cuestionan ciertos puntales del paradigma establecido. No es que se trate de auténtica crítica científica, ya que los provocadores tampoco están dispuestos a entrar en razón. Se trata, simplemente, de que ven el mundo desde un nuevo paradigma. ¿Cómo dieron con él? La presión de la evidencia acumulada, y el hecho de que la explicación de ésta careciera de precisión científica, sencillez y claridad al aplicarle los criterios del viejo paradigma, acabaron por resultarles insoportables. (No está mal, aunque resulta un poco difícil de entender cómo se puede sucumbir a la presión de una evidencia ante la cual se es, según esta misma hipótesis, ciego.) Sea como fuere, empieza una nueva era de ciencia «revolucionaria». El grupo dominante aún trata de hacer ciencia «normal» de acuerdo con el viejo paradigma y lucha contra los revolucionarios con todos los medios a su alcance (interfiriendo publicaciones, apartando a los herejes de cargos académicos, etcétera). Los contestatarios encuentran el medio de publicar sus trabajos, ridiculizan a los viejos carcamales e intentan infiltrarse en instituciones influyentes. El poder explicativo del nuevo paradigma (en sus propios términos, ya que en los del viejo sus explicaciones resultan estrafalarias y poco convincentes) atrae a nuevos adeptos entre los jóvenes científicos aún no comprometidos. Puede también haber deserciones en ambos bandos. Algunos de los vejestorios mueren. Con el tiempo, uno de los bandos gana la batalla. Si triunfan los herejes, se convierten en el nuevo *establishment* científico y defienden su nuevo paradigma con tanta ceguera como defendieron sus antecesores el suyo. Si pierden, se convierten en una nota al pie de página en algún manual de historia de la ciencia. En cualquier caso, la ciencia vuelve a ser «normal».

La concepción de Kuhn del proceso científico le parece natural a mucha gente. Parece explicar los reiterados y desconcertantes cambios que la ciencia ha impuesto al pensamiento moderno, en términos de atributos e impulsos humanos generales y que nos resultan a todos familiares: ideas y

prejuicios arraigados, ceguera ante cualquier evidencia de los propios errores, supresión de la disidencia en defensa de intereses creados, anhelo de estabilidad y rechazo de todo aquello que pueda perturbarla. En el extremo opuesto se encuentran la rebeldía de la juventud, la búsqueda de la novedad, la alegría de romper tabúes y la lucha por el poder. Otro atractivo de las ideas de Kuhn reside en que baja a los científicos de su pedestal. No pueden ya proclamarse nobles buscadores de la verdad que utilizan métodos racionales de conjetura, crítica y comprobación experimental para resolver problemas y crear explicaciones cada vez mejores del mundo. Kuhn nos asegura que se trata tan sólo de equipos rivales enzarzados en luchas interminables por el control del territorio.

La idea de paradigma, en sí, es indiscutible. Observamos y entendemos el mundo a través de una serie de teorías que conforman un paradigma. Pero Kuhn se equivoca al afirmar que defender un paradigma significa no poder ver los méritos de otro, o ser incapaz de cambiar de paradigma o de comprender dos paradigmas a la vez. (Para un estudio más amplio de las implicaciones de este error, véase *The Myth of the Framework*, de Karl Popper.) Sin duda, siempre existirá el peligro de desestimar o, simplemente, no percibir la capacidad explicativa de una nueva teoría fundamental al evaluarla desde el interior del marco conceptual de otra anterior, pero es tan sólo un peligro, y podemos evitarlo poniendo el suficiente cuidado y obrando con la máxima integridad intelectual.

Es también cierto que las personas —incluyendo a los científicos, y, especialmente, a aquellos que ocupan puestos que confieren poder— tendemos a aferramos al modo dominante de hacer las cosas y a recelar de las nuevas ideas cuando nos sentimos a gusto con las existentes. Nadie puede pretender que los científicos sean uniforme y escrupulosamente racionales al juzgar nuevas ideas. La lealtad injustificada a los paradigmas es, sin duda, causa de controversias en ciencia, como en otros muchos campos. Sin embargo, considerada como una descripción o un análisis del proceso científico, la teoría de Kuhn adolece de un defecto fundamental. Explica la *sucesión* de un paradigma por otro en términos sociológicos o psicológicos, en vez de relacionarla básicamente con el mérito objetivo de las explicaciones rivales. Y si no se entiende la ciencia como búsqueda de explicaciones, el hecho de que descubra sucesivamente nuevas explicaciones, cada

una de ellas mejor que la anterior desde un punto de vista objetivo, resulta inexplicable.

De ahí que Kuhn se vea obligado a negar de plano que haya una mejora objetiva en las sucesivas explicaciones, o que tal mejora resulte posible, incluso en principio: «... hay [un paso] que muchos filósofos de la ciencia están dispuestos a dar, pero que rechazo. Consiste en comparar las teorías como si fuesen representaciones de la naturaleza, como si fuesen enunciados acerca de "lo que existe realmente en ella". Aun dando por sentado que ninguna de las dos teorías que se enfrentan en un momento dado de la historia es del todo cierta, buscan, sin embargo, algún sentido en que la más moderna sea una mejor aproximación a la verdad. No creo que sea posible conseguirlo» (Lakatos y Musgrave, eds., *Criticism and the Growth of Knowledge*, página 265).

Así pues, el desarrollo del conocimiento científico objetivo no puede explicarse en el esquema de Kuhn. De nada sirve propugnar que las sucesivas explicaciones son mejores únicamente en términos de su correspondiente paradigma. Existen diferencias objetivas. Podemos volar, a pesar de que durante la mayor parte de la historia de la humanidad sólo se podía soñar con ello. Los antiguos no habrían negado la capacidad de volar de nuestras aeronaves simplemente porque, de acuerdo con su paradigma, les resultara imposible comprender su funcionamiento. La razón por la que podemos volar es que comprendemos «lo que existe realmente en la naturaleza» lo bastante bien para poder construir aeronaves. La razón por la que los antiguos no podían hacerlo, es que su comprensión era objetivamente inferior a la nuestra.

Si aplicamos a la realidad del progreso científico objetivo la teoría de Kuhn, el resultado es que todas las innovaciones fundamentales han sido obra de un puñado de genios iconoclastas. El resto de la comunidad científica no deja de tener cierta utilidad, pero, en materias de importancia, se limita a entorpecer el desarrollo del conocimiento. Esta visión romántica (frecuentemente expuesta con independencia de las ideas de Kuhn) no se corresponde tampoco con la realidad. Han existido, sin duda, genios que, en solitario, han revolucionado ciencias enteras. Algunos han sido ya mencionados en este libro, como Galileo, Newton, Faraday, Darwin, Einstein, Gódel y Turing. Pero, en conjunto, consiguieron investigar, publicar sus trabajos y alcanzar el reconocimiento general *a pesar de* la inevitable

oposición de los inmovilistas y los conformistas. (Galileo fue, ciertamente, acallado, pero no por científicos rivales.) Si bien muchos de ellos se enfrentaron a una oposición irracional, ninguno ejemplificó con su carrera el estereotipo «iconoclasta contra el *establishment* científico». La mayoría de ellos se benefició de sus interacciones con científicos partidarios del paradigma anterior y sacó provecho de ellas.

Me he encontrado en ocasiones en el lado minoritario de controversias científicas fundamentales, pero nunca me he tropezado con una situación parecida a la que expone Kuhn. Por supuesto, y como ya he señalado, la mayor parte de la comunidad científica no está siempre lo abierta a la crítica que sería de desear; sin embargo, la fidelidad de su adhesión a la «adecuada práctica científica» durante el proceso de investigación resulta, francamente, notable. Basta asistir a un seminario dedicado a un problema fundamental de alguna de las ciencias experimentales para darse cuenta de hasta qué punto difiere el comportamiento *como investigadores* de los participantes del comportamiento humano habitual. Dirige el seminario un conocido catedrático, figura destacada de la especialidad. La sala está repleta de científicos de todos los niveles, desde estudiantes acabados de licenciar que entraron en contacto con el campo concreto del seminario hace tan sólo unas semanas hasta profesores cuyo prestigio rivaliza con el del director. La jerarquía científica es una compleja estructura de poder en la que las carreras de las personas, su influencia y su reputación están continuamente en juego, tanto o más que en cualquier gobierno o consejo de administración. Sin embargo, durante el desarrollo del seminario le resultaría difícil a un observador distinguir las categorías de los participantes. Un estudiante recién licenciado plantea una duda: «¿Se desprende realmente su tercera ecuación de la segunda? El término que omitió no me parece despreciable.» El profesor está seguro de que el término *es* despreciable y el estudiante comete un error de apreciación, en el que alguien con más experiencia no habría caído. ¿Qué ocurre entonces?

En una situación análoga, es probable que un alto ejecutivo cuya opinión en una cuestión empresarial fuese contradecida por un empleado acabado de contratar le contestase: «Mire, he resuelto más problemas de esta clase que pelos tiene usted en la cabeza. Si le digo que es así, es que es así.» Un político veterano podría contestar a la crítica de algún oscuro, pero ambicioso,

militante de base del partido algo así: «¿Se puede saber de qué lado estás?» Incluso nuestro profesor, *fuera del contexto del seminario* (impartiendo una clase, por ejemplo) podría contestar despectivamente: «¡Más le valdría estudiar y no malgastar su tiempo ni el nuestro!» Pero en el seminario responder así a una crítica ocasionaría una verdadera conmoción en la sala. Los participantes mirarían hacia otro lado y simularían estar absortos en sus notas. Habría sonrisas forzadas y miradas de reojo. Todos estarían sorprendidos ante una actitud tan escandalosamente inadecuada. En situaciones así, recurrir al principio de autoridad (al menos, de manera explícita) está, simplemente, fuera de lugar, incluso cuando la persona más destacada de las presentes se dirige a la más insignificante.

De modo que el profesor se toma muy en serio el comentario del estudiante y responde con una argumentación concisa, pero adecuada, en defensa de su controvertida ecuación. El profesor trata de no mostrar irritación ante las críticas, por insignificante que sea quien las haga. La *mayoría* de las preguntas que se le dirijan contendrán críticas que, de ser válidas, podrían monoscabar o destruir el trabajo de toda una vida de investigación. Pero someter las verdades comúnmente aceptadas a las críticas más diversas y virulentas forma parte de los objetivos del seminario. Todos los presentes dan por sentado que la verdad no resulta obvia y que lo obvio no tiene por qué ser verdad, que las ideas deben ser aceptadas o rechazadas por su contenido y no por su origen, que las mentes más privilegiadas se pueden equivocar con facilidad y que la objeción en apariencia más trivial puede ser la clave de un gran descubrimiento.

Así pues, los participantes en el seminario, mientras estén ocupados en temas de ciencia, se comportarán en gran medida con racionalidad científica. Pero... termina el seminario. Acompañemos al grupo al comedor. Inmediatamente se restablece el comportamiento social humano normal. El profesor es tratado con deferencia y ocupa en la mesa el lugar central entre los de su rango. Unos pocos elegidos de rangos inferiores reciben el privilegio de ocupar los sitios próximos. La conversación trata del tiempo, de temas comunes y (especialmente) de política universitaria. En la conversación reaparecerán el dogmatismo y los prejuicios, el orgullo y la lealtad, las amenazas y las adulaciones características de cualquier interacción humana, en circunstancias similares. Pero si la conversación deriva hacia los temas objeto del

seminario, los científicos vuelven inmediatamente a comportarse como tales. Se buscan explicaciones, imperan la evidencia y el razonamiento, y el rango deja de ser relevante para el desarrollo del debate. Ésta ha sido, al menos, mi experiencia personal en todos los campos en que he trabajado.

Si bien la historia de la teoría cuántica proporciona múltiples ejemplos de científicos que se aferraron irracionalmente a lo que podríamos denominar «paradigmas», resultaría difícil encontrar una refutación más clara de la teoría de Kuhn de la *sucesión* de paradigmas. El descubrimiento de la teoría cuántica constituyó, sin duda, una revolución conceptual, quizás la mayor después de Galileo, y hubo «viejos carcamales» que nunca la aceptaron. Pero las principales figuras de la física, incluyendo casi todas las que se podrían considerar miembros del *establishment* científico, se mostraron inmediatamente dispuestas a abandonar el paradigma clásico. Pronto se difundió el convencimiento de que la nueva teoría exigía una ruptura radical con el concepto clásico de la estructura de la realidad. El único debate era cuál debía ser la nueva concepción. Transcurrido un tiempo, se estableció una nueva ortodoxia por el físico Niels Bohr y su «escuela de Copenhague». Esta nueva ortodoxia nunca llegó a ser tan ampliamente aceptada en cuanto *descripción de la realidad* para poder ser considerada un paradigma, aunque se adhirió a ella la mayoría de los físicos (Einstein constituyó una notable excepción). Es curioso que esta escuela no se basaba en la proposición de que la nueva teoría cuántica fuese cierta, sino que, al contrario, dependía crucialmente de que fuese, al menos en su fomulación corriente, ¡falsa! Según la «interpretación de Copenhague», las ecuaciones de la teoría cuántica sólo son de aplicación a los aspectos no observados de la realidad física. Cuando tiene lugar la observación, se da otra clase de proceso, que incluye una interacción directa entre la conciencia humana y la física subatómica. Un determinado estado de la conciencia se convierte en real; el resto pasan a ser meras posibilidades. La interpretación de Copenhague sólo especificó este proceso de manera superficial; dejó para más adelante su descripción más profunda, suponiendo que llegara a estar algún día al alcance de la comprensión humana. En cuanto a los acontecimientos no observados que se interpolaban entre las observaciones conscientes, ¡«no se permitía preguntar» sobre ellos! Resulta incomprensible que los físicos —incluso durante el apogeo del positivismo y el instrumentalismo—

aceptaran semejante montaje insustancial como la versión ortodoxa de una teoría fundamental. No necesitamos ocuparnos aquí de los arcaicos detalles de la interpretación de Copenhague, ya que su motivación principal consistía en evitar la conclusión de que la realidad es polivalente y, por ello, incompatible con cualquier explicación genuina de los fenómenos cuánticos.

Unos veinte años después, Hugh Everett, que a la sazón preparaba su doctorado en Princeton bajo la tutela del eminente físico John Archibald Wheeler, fue el primero en establecer las implicaciones multiversales de la teoría cuántica. Wheeler las rechazó. Estaba (y sigue estándolo) convencido de que la visión de Bohr, si bien incompleta, era la base para la explicación correcta. ¿Se comportó, sin embargo, como sería de esperar según el estereotipo de Kuhn? ¿Trató de suprimir las ideas heréticas de su doctorando? Bien al contrario, Wheeler temió que las ideas de Everett no fuesen lo suficientemente valoradas, así que escribió una breve introducción que acompañaba al trabajo de su alumno, y ambos escritos aparecieron en páginas consecutivas de *Reviews of Modem Physics.* El trabajo de Wheeler explicaba y defendía el de Everett tan eficazmente, que muchos lectores llegaron a la conclusión de que ambos eran responsables de su contenido. En consecuencia, y para consternación de Wheeler, la teoría multiversal fue erróneamente conocida como la «teoría de Everett-Wheeler» durante muchos años.

La ejemplar fidelidad de Wheeler a la racionalidad científica puede parecer extrema, pero no es, ni mucho menos, única. En este sentido, debo mencionar a Bryce DeWitt, otro eminente físico que se opuso inicialmente a Everett. En un histórico intercambio de correspondencia, DeWitt formuló una serie de objeciones técnicas detalladas a la teoría de Everett que fueron refutadas una tras otra por éste. DeWitt dio por terminada la discusión con una nota informal en la que señalaba que, simplemente, no notaba que se «trocease» en múltiples copias distintas de sí mismo cada vez que tenía que tomar una decisión. La respuesta de Everett recuerda la disputa entre Galileo y la Inquisición: «¿Nota que la Tierra se mueve?», le preguntó, aludiendo a que la teoría cuántica *explica precisamente por qué* no notamos esta multiplicidad, del mismo modo que la teoría de Galileo sobre la inercia explica por qué no notamos el movimiento de la Tierra. DeWitt se rindió.

No obstante, el descubrimiento de Everett no fue ampliamente aceptado. Por desgracia, la mayoría de los físicos de

la generación situada entre la interpretación de Copenhague y Everett habían abandonado la idea de la explicación en el ámbito de la teoría cuántica. Como he dicho, eran tiempos de apogeo del positivismo en la filosofía de la ciencia. El rechazo (o la incomprensión) ante la interpretación de Copenhague, sumado a lo que podríamos denominar *instrumentalismo pragmático*, se convirtió (y sigue siéndolo) en la actitud típica del físico hacia la teoría más profunda conocida sobre la realidad. Si el instrumentalismo es la doctrina que afirma que las explicaciones son inútiles, puesto que una teoría es tan sólo un «instrumento» para formular predicciones, el instrumentalismo pragmático, en sí, es la práctica de utilizar las teorías sin saber lo que significan ni preocuparse por ello. A este respecto, el pesimismo de Kuhn sobre la racionalidad científica quedaba confirmado, pero no su teoría sobre cómo unos paradigmas sustituyen a otros. En cierto sentido, el instrumentalismo pragmático se convirtió por sí mismo en un «paradigma», adoptado por los físicos para reemplazar la idea clásica de una realidad objetiva. Pero, ciertamente, ¡no es la clase de paradigma que permite comprender el mundo! En cualquier caso, y sea cual fuere la investigación que estuviesen haciendo los físicos, no concebían precisamente el mundo según el paradigma de la física clásica, epítome —entre otras cosas— del realismo y el determinismo objetivos. La mayor parte de ellos lo abandonaron tan pronto como fue formulada la teoría cuántica, a pesar de que había dominado todos los aspectos de la ciencia, incuestionado, desde que Galileo ganó su pleito intelectual contra la Inquisición tres siglos antes.

El instrumentalismo pragmático ha sido posible porque, en la mayoría de las ramas de la física, la teoría cuántica no se aplica en su aspecto explicativo. Se utiliza indirectamente para la comprobación de otras teorías, y tan sólo se necesitan sus predicciones. Generaciones de físicos han creído, pues, suficiente considerar a los procesos de interferencia —por ejemplo, los que tienen lugar durante una milbillonésima de segundo cuando colisionan dos partículas elementales—, como una «caja negra»: preparan unos datos de entrada y observan otros de salida. Utilizan las ecuaciones de la teoría cuántica para predecir la una de la otra, pero no saben, ni les preocupa, *cómo se produce el resultado* consecuencia de la entrada. No obstante, hay dos ramas de la física en las que esta actitud resulta inviable porque se ocupan intrínsecamente del funcionamiento interno del objeto

cuantomécanico. Son la teoría cuántica de la calculabilidad y la cosmología cuántica (la teoría cuántica de la realidad física como un todo). ¡Sería una pobre «teoría de la calculabilidad», realmente, aquella que nunca se planteara cómo se obtienen los resultados a partir de los datos iniciales! Por lo que respecta a la cosmología cuántica, no podemos preparar datos iniciales al principio del universo ni medir resultados al final. Su funcionamiento interno es todo lo que hay. Por esta razón, la teoría cuántica es utilizada en su forma plena y multiversal por la mayoría de los investigadores que trabajan en ambos campos.

Así pues, la historia de Everett es, efectivamente, la de un joven investigador con ganas de innovar, que desafía al consenso dominante y es ignorado por casi todo el mundo hasta que, décadas después, su concepción se convierte poco a poco en el nuevo consenso. Pero la base de la innovación de Everett no consistió en proclamar que la teoría dominante es falsa, sino cierta. El problema era que los científicos, lejos de pensar únicamente en términos de dicha teoría dominante, rehusaban hacerlo y la utilizaban tan sólo de manera instrumental, a pesar de que habían abandonado su anterior paradigma explicativo, la física clásica, sin mayores reparos tan pronto como dispusieron de una teoría mejor.

Algo parecido a este extraño fenómeno ha ocurrido con las otras tres teorías que ofrecen las principales vías de explicación de la estructura de la realidad: las teorías de la calculabilidad, la evolución y el conocimiento. En todos los casos la teoría dominante en la actualidad, si bien ha desplazado a su predecesora y a otras rivales en el sentido de que es aplicada de manera pragmática de modo habitual, no ha llegado a convertirse en el nuevo «paradigma». Es decir, no ha sido admitida por los que trabajan en ese campo como explicación fundamental de la realidad.

El principio de Turing, por ejemplo, rara vez ha sido cuestionado como verdad pragmática, al menos en sus formulaciones más sencillas (por ejemplo, que un ordenador universal podría reproducir cualquier entorno físicamente posible). Las críticas de Roger Penrose constituyen una insólita excepción, ya que no ignora que contradecir el principio de Turing implica considerar teorías radicalmente nuevas tanto en física como en epistemología, al igual que en biología. Pero nadie, ni siquiera Penrose, ha propuesto aún una teoría viable que se oponga al

principio de Turing, de modo que sigue siendo la teoría fundamental dominante por lo que se refiere a la calculabilidad. Sin embargo, la proposición de que la *inteligencia artificial* es, en principio, posible, que se desprende por simple lógica de dicha teoría dominante, no se da en absoluto por sentada. (Una inteligencia artificial es un programa de ordenador que posee propiedades de la mente humana, como inteligencia, conciencia, libre albedrío y emociones, pero que se ejecuta mediante un soporte distinto del cerebro humano.) La posibilidad de que exista la inteligencia artificial es agriamente contestada por eminentes filósofos (incluyendo, por desgracia, al propio Popper), científicos y matemáticos, así como por, al menos, un destacado experto en el campo de la informática. Sin embargo, pocos de ellos parecen darse cuenta de que al hacerlo contradicen el principio básico de una disciplina fundamental sin (a diferencia de Penrose) proponer alternativas. Es como si negasen la posibilidad de viajar a Marte sin darse cuenta de que nuestras más modernas teorías de ingeniería y física nos dicen que es posible. Infringen así un principio básico de la racionalidad: el de que las buenas explicaciones no deben descartarse a la ligera.

No son sólo los oponentes de la inteligencia artificial quienes no han incorporado el principio de Turing a su paradigma. Muy pocos científicos lo han hecho. Buena prueba de ello es que pasaron cuatro décadas desde que fue propuesto hasta que alguien investigó sus implicaciones para la física, y otra década hasta que fue descubierta la teoría cuántica. Los científicos aceptaban y usaban el principio pragmáticamente en sus ordenadores, pero no lo integraban en su concepción general del mundo.

La epistemología de Popper se ha convertido, a todos los efectos pragmáticos, en la teoría dominante por lo que respecta a la naturaleza y el desarrollo del conocimiento científico. Tanto si se trata de definir las reglas de experimentación aceptadas por los teóricos de cualquier campo como «evidencia científica» como de establecer los criterios exigidos por las revistas científicas para publicar trabajos o los seguidos por los médicos para elegir entre tratamientos rivales, el vocabulario fundamental sigue siendo el que utilizaría Popper: prueba experimental, exposición a la crítica, explicación teorética y reconocimiento de la falibilidad de los procedimientos experimentales. En los textos populares de divulgación científica, las teorías tienden a ser presentadas más

como audaces conjeturas que como inferencias extraídas de datos acumulados, y la diferencia entre la ciencia y (digamos) la astrología es explicada correctamente en términos de posibilidad de comprobación más que de grado de confirmación. En los laboratorios de las escuelas, la «formulación y comprobación de hipótesis» está a la orden del día. No se espera ya que los alumnos «aprendan por la experimentación» en el sentido en que lo hicimos mis contemporáneos y yo: se nos proporcionaba un equipo y se nos daban las instrucciones sobre lo que hacer con él, pero sin ser informados de la teoría a la que se deberían conformar los resultados que obtuviésemos. Se esperaba de nosotros que la indujéramos.

A pesar de ser la teoría predominante en este sentido, la epistemología popperiana forma parte de la visión del mundo de muy pocas personas. La popularidad de la teoría de Kuhn sobre la sucesión Je paradigmas constituye un buen ejemplo de ello. Más seriamente, son escasos los filósofos que se muestran de acuerdo con la afirmación de Popper de que ya no existe el «problema de la inducción» porque, de hecho, ya no obtenemos o justificamos las teorías a partir de observaciones, sino que procedemos mediante conjeturas explicativas y refutaciones. No es que la mayoría de los filósofos sean inductivistas, discrepen seriamente de la descripción que hace Popper del método científico que propugna o crean que las teorías científicas son, en realidad, poco sólidas a causa de su status conjetural. Lo que ocurre, simplemente, es que no aceptan la *explicación* de Popper sobre cómo funciona todo ello. He aquí, de nuevo, un eco del caso de Everett. El punto de vista mayoritario es que hay un problema filosófico fundamental con la metodología popperiana, aunque la ciencia la haya aplicado siempre allí donde podía hacerlo con éxito. La herética innovación de Popper toma la forma de una tesis que sostiene que la metodología ha sido siempre válida.

La teoría de Darwin de la evolución es también la dominante en su campo, en el sentido de que nadie duda de que la evolución mediante la selección natural, al actuar sobre poblaciones con variaciones aleatorias, es el «origen de las especies» y de la adaptación biológica en general. Ningún biólogo o filósofo serio atribuiría el origen de las especies a la creación divina o a la evolución tal como la preconizaba Lamarck. (El lamarquismo, teoría evolutiva que precedió al darwinismo, era análogo al inductivismo. Atribuía las adaptaciones biológicas a la herencia de

características que el organismo se había visto obligado a adquirir durante su vida como consecuencia de presiones externas.) Sin embargo, al igual que ocurre con las otras tres vías, las objeciones al darwinismo puro *como explicación* para los fenómenos de la biosfera son numerosas y están muy extendidas. Una fuente de objeciones es la cuestión de si la historia de la biosfera ha durado el tiempo suficiente para que la tremenda complejidad presente haya evolucionado únicamente por selección natural. Ninguna teoría rival viable ha sido expuesta en apoyo de esas objeciones, si se exceptúa la idea, de la que los astrónomos Fred Hoyle y Chandra Wickramasinghe son recientes proponentes, de que las complejas moléculas en que se basa la vida hubiesen podido originarse en el espacio exterior. Pero el objetivo de dichas objeciones, más que contradecir el modelo darwiniano, es recordar que quedan por explicar muchas cosas fundamentales acerca del origen de las adaptaciones que observamos en la biosfera.

El darwinismo ha sido también criticado por caer en un círculo vicioso, ya que invoca «la supervivencia de los más aptos» como explicación, mientras que los «más aptos» son definidos de modo retrospectivo por ser, precisamente, los que sobrevivieron. Por otra parte, y en términos de una definición independiente de «más apto», la idea de que la evolución «favorece a los más aptos» parece estar en contradicción con los hechos. Por ejemplo, la definición más intuitiva de aptitud biológica sería la «aptitud de una especie para la supervivencia en un nicho determinado», en el sentido de que un tigre podría ser considerado la máquina óptima para ocupar el nicho ecológico destinado a los tigres. Ejemplos típicos contrarios a esa «supervivencia de los más aptos» son las adaptaciones, como la cola del pavo real, que parecen causar que el organismo sea mucho *menos* apto para explotar su nicho. Tales objeciones parecen poner en entredicho la capacidad de la teoría de Darwin para cumplir su propósito original, que no era otro que explicar hasta qué punto el aparente «diseño» (adaptaciones) de los organismos vivos podría ser consecuencia de la acción de leyes físicas «ciegas» sobre la materia inanimada, sin la intervención intencionada de ningún «Diseñador».

La innovadora proposición de Richard Dawkins, expuesta en sus obras *El relojero ciego y El gen egoísta*, es otra argumentación en favor de que la teoría dominante es, después de todo, cierta. Aduce que ninguna de las objeciones corrientes al modelo darwiniano esencial resulta tener, si se examina a fondo,

fundamento alguno. En otras palabras, Dawkins proclama que la teoría de la evolución de Darwin proporciona una explicación completa sobre el origen de las adaptaciones biológicas. Dawkins elabora una versión moderna de la teoría de Darwin, conocida como teoría de los replicantes. El replicante que mejor consiga hacerse replicar en un determinado entorno, desplazará con el tiempo a todas sus variantes, ya que, por definición, éstas no consiguen hacerse replicar tan bien. No es la variante más apta de *especie* la que sobrevive (Darwin no se dio cuenta de ello), sino la variante más apta de *gen.* Consecuencia de este hecho es que, algunas veces, un gen puede desplazar a genes variantes (como genes que originarían colas de pavo real menos aparatosas) por medios (tales como la selección sexual) que no promueven especialmente el bien de la especie o el individuo. Pero toda evolución promueve el «bien» (la replicación) de los genes mejores replicantes, de ahí el término «gen egoísta». Dawkins se enfrenta detalladamente a todas las objeciones a la teoría de Darwin y demuestra que ésta, interpretada del modo correcto, no tiene ninguno de los fallos que le atribuyen sus detractores y explica realmente el origen de las adaptaciones.

La versión de Dawkins del darwinismo se ha convertido en la teoría dominante de la evolución en sentido pragmático, si bien no constituye aún en modo alguno el *paradigma* dominante. Muchos biólogos y filósofos se sienten todavía inquietos por la sensación de que hay algún vacío fundamental en la explicación que ofrece. Y, de hecho, en el mismo sentido en que la teoría de Kuhn sobre las «revoluciones científicas» desafía la imagen popperiana de la ciencia, hay una teoría evolutiva alternativa que desafía la imagen de Dawkins de la evolución. Se trata de la teoría del *equilibrio intermitente*, que postula que la evolución sigue una pauta de súbitos brotes separados por períodos más o menos largos sin cambios selectivos. Esta teoría podría incluso resultar factualmente cierta, pero no contradice la teoría del «gen egoísta» más de lo que pueda contradecir a la epistemología popperiana la proposición de que las revoluciones conceptuales no ocurren todos los días, o la de que los científicos se resisten a menudo a la innovación fundamental. Pero, como sucede con la teoría de Kuhn, el modo en que han sido presentados el equilibrio intermitente y otras variantes de escenario evolutivo como solución de algún supuesto problema que ha pasado inadvertido para la teoría dominante revela hasta qué extremo el poder explicativo de la

teoría de Dawkins tiene aún que ser asimilado.

El hecho de que la teoría dominante haya sido rechazada de modo general como explicación sin que hayan surgido, a pesar de ello, explicaciones rivales serias, ha tenido una desafortunada consecuencia para las cuatro vías. Se trata de que los proponentes de las cuatro teorías dominantes —Popper, Turing, Everett y Dawkins—, así como sus seguidores, se han visto obligados a estar constantemente a la defensiva frente a teorías obsoletas. El debate entre Popper y la mayoría de sus críticos se centró (como he expuesto en los capítulos 3 y 7), en efecto, en el problema de la inducción. Turing se pasó los últimos años de su vida defendiendo la proposición de que el cerebro humano no opera por medios sobrenaturales. Everett abandonó la investigación científica al no conseguir hacer progresos, y durante varios años la teoría del multiverso fue defendida casi en solitario por Bryce DeWitt, hasta que los avances experimentados por la cosmologia cuántica en los años setenta conllevaron su aceptación pragmática en dicho campo. Pero los oponentes de la teoría del multiverso *como explicación* raramente han propuesto explicaciones alternativas. (La de David Bohm, mencionada en el capítulo 4, es una excepción.) En vez de ello, como señaló en cierta ocasión el cosmólogo Dennis Sciama, «cuando se trata de la interpretación de la mecánica cuántica, el nivel de la argumentación desciende súbitamente a cero». Los defensores de la teoría del multiverso se enfrentan a menudo a melancólicas y desafiantes, pero incoherentes, apelaciones a la interpretación de Copenhague, en la que, sin embargo, casi nadie cree ya. Y, finalmente, Dawkins se ha convertido, hasta cierto punto, en el defensor público de la racionalidad científica contra el *creacionismo*, aunque parezca mentira, y, más generalmente, contra la concepción precientífica del mundo, absoleta desde Galileo. Lo más frustrante de todo ello es que mientras quienes proponen las mejores teorías de que disponemos sobre la estructura de la realidad tengan que malgastar sus energías en inútiles refutaciones reiteradas de teorías que se sabe desde hace tiempo que son falsas, el estado de nuestro conocimiento profundo no podrá mejorar. Tanto Turing como Everett habrían podido descubrir perfectamente la teoría cuántica de la calculabilidad. Popper habría podido elaborar la teoría de la explicación científica. (En honor a la verdad, hay que reconocer que comprendió y estableció algunas conexiones entre su epistemología y la teoría de la evolución.) Dawkins podría, por

ejemplo, trabajar en su nueva teoría de la evolución de las ideas replicantes (memes).

La teoría unificada de la estructura de la realidad que constituye el tema de este libro es, simplemente, en su nivel más sencillo, la combinación de las cuatro teorías fundamentales dominantes en sus respectivos campos. En este sentido, es también la «teoría dominante» en estos cuatro campos, considerados como un todo. Algunas de las conexiones entre ellas están incluso ampliamente reconocidas. Mi tesis, por lo tanto, adopta también la postura de que «¡después de todo, la teoría dominante es cierta!». No sólo abogo por tomar cada una de las cuatro teorías en serio como explicación de su materia correspondiente, sino que defiendo que, consideradas en conjunto, proporcionan un nuevo nivel de explicación de la estructura unificada de la realidad.

He argumentado también que cada una de las cuatro vías puede ser entendida con independencia de las demás. Ésta es quizás la clave de que no hayan sido creídas. Las cuatro comparten una poca atractiva propiedad que ha sido variadamente criticada como «idealizada e irreal», «estrecha» e «inocente», así como «fría», «mecanicista» y «falta humanidad». Creo que, en parte, resulta comprensible el sentimiento visceral de rechazo que se oculta tras estos calificativos. Por ejemplo, algunos de quienes niegan la posibilidad de la inteligencia artificial —lo que les conduce a negar que el cerebro sea un objeto físico—, sólo tratan de expresar una crítica mucho más razonable: que la explicación que da Turing de la calculabilidad parece no dejar espacio, incluso en principio, para ninguna explicación futura, *en términos físicos*, de atributos mentales tales como la conciencia y el libre albedrío. No es, pues, suficiente que los entusiastas de la inteligencia artificial les respondan abruptamente que el principio de Turing garantiza que un ordenador puede hacer todo lo que sea capaz de hacer un cerebro. Ello es, por supuesto, cierto, pero es una respuesta en términos de predicción, mientras que el problema es de explicación. Hay, pues, un *vacío explicativo.*

No creo que este vacío pueda ser llenado sin recurrir a las otras tres vías. Ahora bien, como ya he dicho, mi opinión es que el cerebro es un ordenador clásico y no cuántico, de modo que no espero que la explicación de la conciencia consista en que ésta es alguna clase de fenómeno cuantocalculatorio. Sin embargo, sí que creo que la unificación de la calculabilidad y la física cuántica —y,

probablemente, de las cuatro vías— será esencial para los avances *filosóficos* fundamentales a partir de los cuales deberá llegar algún día la comprensión de la conciencia. Si al lector esto le resulta paradójico, permítame esbozar una analogía con un problema parecido de una época anterior. ¿Qué es la vida? Darwin resolvió esta cuestión. La esencia de la solución fue la idea de que el diseño intrincado y, al menos en apariencia, dirigido hacia un objetivo de los organismos vivos no está incorporado *ab initio* a la realidad, sino que es una consecuencia emergente de la acción de las leyes de la física. Ni éstas ni ningún creador han dictado específicamente el aspecto de los elefantes y los pavos reales. No hacen referencia a los resultados, y mucho menos a los emergentes, sino que determinan, simplemente, las reglas según las cuales deben interactuar los átomos y las demás partículas. Ahora bien, este concepto de que una ley de la naturaleza es un conjunto de leyes del movimiento es bastante reciente. Debe, en mi opinión, ser atribuido específicamente a Galileo, así como, y en cierta medida, a Newton. El concepto anterior de ley de la naturaleza era el de una regla que describe *lo que sucede.* Encontramos un buen ejemplo de ello en el contraste entre las leyes del movimiento planetario de Johannes Kepler, que describían cómo se mueven los planetas en órbitas elípticas, y las de Newton, que constituyen leyes en el moderno sentido de la palabra. No hacen mención de las elipses, pero reproducen (y corrigen) las predicciones de Kepler en las condiciones adecuadas. Nadie habría podido explicar qué es la vida mediante el concepto de «ley de física» de Kepler, puesto que habría buscado inútilmente una ley que determinara a los elefantes, del mismo modo que las leyes de Kepler determinaban a las elipses. Darwin, sin embargo, fue capaz de preguntarse por qué unas leyes de la naturaleza que no dictaban la existencia de los elefantes eran, no obstante, capaces de producirlos, del mismo modo que las de Newton producían elipses. Si bien Darwin no hizo uso específico de ninguna de las leyes de Newton, su descubrimiento habría sido inconcebible sin la concepción del mundo inmanente en dichas leyes. Éste es el sentido en el que espero que la solución al problema «¿Qué es la conciencia?» se base en la teoría cuántica. No invocará procesos cuantomecánicos específicos, pero dependerá crucialmente de una imagen del mundo cuantomecánica y, muy especialmente, multiversal.

¿De qué pruebas dispongo? He presentado ya algunas en el

capítulo 8, en el que analicé la concepción multiversal del conocimiento. Si bien no sabemos qué es la conciencia, ésta se halla clara e íntimamente relacionada con el desarrollo y la representación del conocimiento en el interior del cerebro. No parece, pues, probable que podamos estar en condiciones de explicar lo que es, como proceso físico, sin haber explicado antes el conocimiento en términos físicos. Semejante explicación se le ha escurrido siempre de entre los dedos a la teoría clásica de la calculabilidad, pero, como ya he dicho, ahora disponemos de una buena base para conseguirlo gracias a la teoría cuántica: el conocimiento puede ser entendido como una complejidad que se extiende a través de múltiples universos.

Otro atributo mental asociado en cierto modo con la conciencia es el libre albedrío, que resulta igualmente difícil de entender a partir de la imagen clásica del mundo. La dificultad de reconciliar el libre albedrío con la física es atribuida frecuentemente al determinismo, pero éste no es el culpable. Lo es (como he explicado en el capítulo 11) el espacio-tiempo clásico. En éste *algo* me sucede en cada momento particular de mi futuro. Aun siendo impredecible, lo que me va a ocurrir está ya ahí, en la correspondiente sección del espacio-tiempo. Carece de sentido hablar de que yo «cambie» lo que está en esa sección. El espacio-tiempo no cambia, y, por consiguiente, no podemos, dentro de su física, concebir causas, efectos, futuro abierto o libre albedrío.

Así pues, reemplazar las leyes deterministas del movimiento por otras no deterministas (aleatorias) no ayudaría, ni mucho menos, a resolver el problema del libre albedrío si las leyes continuaban siendo clásicas. La libertad no tiene nada que ver con la aleatoriedad. Valoramos nuestro libre albedrío como la capacidad de expresar, mediante nuestras acciones, lo que somos como individuos. ¿Quién valoraría ser aleatorio? Lo que consideramos nuestras acciones *libres* no son las aleatorias o indeterminadas, sino las que están ampliamente *determinadas* por quienes somos, cómo pensamos y qué está en juego. (Si bien están ampliamente determinadas, pueden ser muy impredecibles por razones de complejidad.)

Consideremos este enunciado típico referido al libre albedrío: «Después de considerarlo cuidadosamente, escojo X; hubiera podido escoger otra cosa; tomé la decisión acertada; se me da bien tomar esta clase de decisiones.» En cualquier imagen clásica del mundo, este enunciado es un galimatías. Desde la perspectiva

del multiverso, tiene una representación física directa, que se muestra en la figura 13.1. (No se trata aquí de *definir* valores estéticos o morales en términos de semejantes representaciones, sino, simplemente, de señalar que, gracias al carácter multiversal de la realidad cuántica, el libre albedrío y los conceptos con él relacionados son ahora compatibles con la física.)

De este modo, el concepto de calculabilidad de Turing parece menos desconectado de los valores humanos, y no constituye ya un obstáculo para la comprensión de atributos como el libre albedrío, a condición de entenderlo en el contexto del multiverso. El mismo ejemplo libera de toda culpa a la propia teoría de Everett. A primera vista, parece que el precio de la comprensión de los fenómenos de interferencia es la creación y la exacerbación de multitud de problemas filosóficos. Pero, en realidad, y como sucede con muchos de los ejemplos que he ofrecido en este libro, ocurre exactamente lo contrario. La utilidad de la teoría del multiverso para contribuir a la solución de problemas filosóficos que existen desde hace mucho tiempo es tan grande, que valdría la pena adoptarla aunque no la apoyase la evidencia física. De hecho, el filósofo David Lewis, en su libro *On the Plurality of'Worlds*, postula la existencia del multiverso por razones únicamente filosóficas.

Volviendo a la teoría de la evolución, puedo, de modo similar, comprender, *hasta cierto punto*, a quienes critican la visión darwiniana de la evolución sobre la base de que parece «poco probable» que adaptaciones tan complejas puedan haber evolucionado en el tiempo que se les atribuyese. Uno de los críticos de Dawkins nos propone que nos mostremos tan sorprendidos ante la biosfera como lo estaríamos si un montón de piezas se convirtiese por sí mismo en un Boeing 747. A primera vista, esta crítica establece una forzada analogía entre, por una parte, miles de millones de años de prueba y error sobre el planeta, y, por otra, el acontecimiento fortuito e instantáneo de una adaptación espontánea. Ello equivaldría a obviar de plano todo

el meollo de la explicación evolutiva. No obstante, ¿es del todo válida la postura exactamente opuesta de Dawkins como explicación? Dawkins nos propone que *no nos sorprendamos* de que hayan ocurrido adaptaciones complejas de modo espontáneo. En otras palabras, proclama que su teoría del «gen egoísta» constituye una explicación completa; no, por supuesto, de las adaptaciones específicas, sino de cómo fue posible que dichas adaptaciones llegaran a existir.

Pero no es una explicación completa. Hay un vacío explicativo, *y*, en esta ocasión, sabemos ya mucho más sobre cómo lo podrían completar las otras tres vías. Hemos visto que el propio hecho de que las variables físicas puedan almacenar información, puedan interactuar entre sí para transferirla y replicarla, y el proceso sea estable, depende en su totalidad de las circunstancias de la teoría cuántica. Y, lo que es más, hemos visto que la existencia de replicantes altamente adaptados depende de la factibilidad física de la generación y la universalidad de la realidad virtual, las cuales, a su vez, pueden ser entendidas como consecuencias de un principio profundo, el principio de Turing, que vincula la física con la teoría de la calculabilidad sin referencia explícita alguna a replicantes, evolución o biología.

Un vacío análogo se da en la epistemología popperiana. Sus críticos se preguntan por qué funciona el método científico, o qué justifica nuestra confianza en las mejores teorías científicas. Ello los lleva a suspirar por un principio de inducción o algo parecido (si bien, como los criptoinductivistas, reconocen habitualmente que tal principio tampoco explicaría ni justificaría nada). La respuesta popperiana de que no existe la justificación, o de que nunca es racional confiar en las teorías, no es una explicación. Popper llegó a decir que «ninguna teoría del conocimiento debería intentar explicar cómo conseguimos explicar las cosas» (*Objective Knowledge*, página 23). Una vez que entendemos que el desarrollo del conocimiento humano es un proceso físico, vemos que no puede ser ilícito explicar cómo y por qué ocurre. La epistemología es una teoría de la física (emergente). Es una teoría factual sobre las circunstancias en que podrá o no crecer una determinada magnitud física (el conocimiento). Las aserciones esenciales de esta teoría son ampliamente aceptadas, pero no hay modo de encontrar ninguna explicación de por qué son ciertas sólo dentro de la teoría del conocimiento *per se.* En este sentido limitado, Popper está en lo cierto. La explicación debe involucrar a la física

cuántica, al principio de Turing y, como el mismo Popper hizo notar, a la teoría de la evolución.

Los proponentes de la teoría dominante, en los cuatro casos, se ven permanentemente forzados a una posición defensiva ante los ataques de sus críticos a esos vacíos explicativos. Ello les obliga a menudo a retraerse al núcleo de su propia vía. «Aquí estoy, y esto es lo que hay», es su respuesta final y un modo de manifestar que se niegan a la autoevidente irracionalidad de abandonar la teoría fundamental y no rebatida de sus campos particulares. Esto sólo consigue hacerles parecer aún más estrechos de miras a ojos de sus críticos, y tiende a generar pesimismo sobre la posibilidad misma de una futura explicación fundamental.

A pesar de todas las muestras de comprensión que he tenido hacia los críticos de las teorías principales, la historia de las cuatro vías nos muestra que algo muy desagradable le ha sucedido a la ciencia fundamental durante la mayor parte del siglo XX. La popularidad del positivismo y de la visión instrumentalista de la ciencia estaba directamente relacionada con la apatía, la falta de confianza en uno mismo y el pesimismo sobre las explicaciones genuinas en una época de vacas gordas en que el prestigio, la utilidad y, por supuesto, la financiación para la investigación fundamental lo eran todo. Hubo, por supuesto, muchas excepciones individuales, entre las que cabe contar a los cuatro héroes de este capítulo. Pero el modo sin precedentes en que sus teorías eran aceptadas e ignoradas a un tiempo habla por sí mismo. No presumo de conocer la explicación de este fenómeno, pero, sea lo que fuera lo que lo causó, parece que está remitiendo últimamente. He señalado una de las posibles causas coadyuvantes, es decir, que, individualmente, las cuatro teorías tienen vacíos explicativos que las pueden hacer parecer estrechas, inhumanas y pesimistas. Sugiero, sin embargo que, tomadas en conjunto como explicación unificada de la estructura de la realidad, esta desafortunada propiedad se invierte. Lejos de negar el libre albedrío, lejos de situar a los valores humanos en un contexto en que pasan a ser triviales e insignificantes, lejos de ser pesimista, constituye una concepción del mundo fundamentalmente optimista, que sitúa a la mente humana en el centro del universo físico y a la explicación y la comprensión en el centro de los propósitos humanos. Espero que no tengamos que perder demasiado tiempo mirando hacia el pasado para defender

esta concepción unificada ante competidores aún por surgir. Éstos no faltarán, sin duda, cuando, tras haber considerado seriamente la teoría unificada de la estructura de la realidad, empecemos a desarrollarla en profundidad.

Es hora de ponerse en marcha.

TERMINOLOGÍA

Paradigma. Conjunto de ideas que sirve a sus adeptos para interpretar todo lo que se relaciona con su experiencia. Según Thomas Kuhn, adherirse a un paradigma impide reconocer los méritos de cualquier otro, por lo que no es posible cambiar de paradigma o comprender dos paradigmas al mismo tiempo.

Interpretación de Copenhague de la mecánica cuántica. Idea que facilita evadirse de las implicaciones de la teoría cuántica en la naturaleza de la realidad. Supone que, en los momentos de observación, el resultado en uno de los universos se convierte en real, y que los demás universos —incluyendo los que contribuyeron al resultado— nunca existieron. De acuerdo con semejante concepción de la realidad, no cabe preguntarse qué ocurre en ella en los intervalos entre las observaciones conscientes.

SUMARIO

La historia intelectual de las teorías fundamentales de las cuatro vías contienen paralelismos notables. Las cuatro han sido aceptadas (para su uso práctico) e ignoradas (como explicaciones de la realidad) a un tiempo. Una razón para ello es que, consideradas individualmente, todas presentan vacíos explicativos y parecen frías y pesimistas. Una visión del mundo basada exclusivamente en cualquiera de ellas sería reduccionista, pero ocurre todo lo contrario cuando se consideran en conjunto como

explicación unificada de la estructura de la realidad.

Y, ahora, ¿qué?

14. ¿CÓMO SE ACABARÁ EL UNIVERSO?

Aunque la historia carezca de significado, se lo podemos atribuir.

Karl Popper

The Open Society and its Enemies, vol. 2, pág. 278

Cuando, en el transcurso de mi investigación sobre los fundamentos de la teoría cuántica, empecé a darme cuenta de los vínculos entre la física cuántica, la calculabilidad y la epistemología, los percibí como una evidencia más de la tendencia histórica de la física a absorber materias que previamente parecían no estar relacionadas con ella. La astronomía, por ejemplo, quedó vinculada a la física terrestre por las leyes de Newton, y, en los siglos siguientes, fue absorbida en gran parte hasta convertirse en la astrofísica. La química empezó a quedar bajo el manto de la física con los descubrimientos de Faraday sobre electroquímica, mientras que la teoría cuántica ha hecho que buena parte de la química básica sea directamente predecible a partir de las leyes de la física. La teoría general de la relatividad de Einstein absorbió a la geometría y sacó tanto a la cosmología como a la teoría del tiempo de su status anterior, sólo filosófico, para convertirlas en ramas plenamente integradas de la física. En las últimas décadas, como he expuesto, la teoría del viaje en el tiempo lo ha sido a su vez.

Así pues, la expectativa de que la física cuántica absorbiera no tan sólo la teoría de la calculabilidad sino, sobre todo, la *teoría de la demostración* (denominada alternativamente «metamatemáticas»), me parecía que evidenciaba dos tendencias. La primera, que el conocimiento humano, en conjunto, seguía adquiriendo la estructura unificada que debería tener para ser comprensible en el sentido total que yo esperaba. Y la segunda, que esa estructura unificada tendría como base una teoría de la física fundamental que se ampliaría y profundizaría sin cesar.

El lector habrá notado que mi opinión ha cambiado en relación al segundo punto. El carácter de la estructura de la

realidad que propongo ahora no se basa únicamente en la física fundamental. Por ejemplo, la teoría cuántica de la calculabilidad no ha sido elaborada derivando principios de calculabilidad sólo de la física cuántica. Incluye el principio de Turing, que, con la denominación de *conjetura de Church-Turing*, constituyó en su momento la base para la teoría de la calculabilidad. Nunca había sido utilizado en física, pero ya he argumentado que, precisamente, sólo como principio de la física puede ser comprendido del modo adecuado. Lo mismo ocurre con el principio de conservación de la energía y las demás leyes de la termodinámica; es decir, es una condición a la que deben conformarse todas las teorías. Pero, a diferencia de las leyes de la física tradicionales, ese principio tiene un carácter emergente, relacionado de manera directa con las propiedades de máquinas complejas y tan sólo por derivación con objetos y procesos subatómicos. (Es discutible que la segunda ley de la termodinámica —el principio del aumento de la entropía— tenga el mismo carácter.)

De modo parecido, si entendemos el *conocimiento* y la *adaptación* como estructuras que se extienden por múltiples universos, podemos esperar que los principios de epistemología y evolución estén expresados en forma de leyes acerca de la estructura del multiverso, es decir, que sean leyes físicas, pero a nivel emergente. Ciertamente, la teoría cuántica de la complejidad no ha alcanzado aún el punto en que podamos expresar, en términos físicos, la proposición de que el conocimiento sólo puede desarrollarse en situaciones que se adapten al patrón popperiano expresado en la figura 3.3. Pero ésta es, precisamente, la clase de proposición que espero ver surgir de la naciente teoría total de la realidad, la teoría unificada explicativa y predictiva de las cuatro vías.

De ahí que la idea de que la física cuántica está absorbiendo a las otras vías deba ser considerada únicamente una estrecha perspectiva de físico, teñida quizás de reduccionismo. En efecto, cada una de las otras tres vías es lo bastante rica para formar la base completa de la concepción del mundo de mucha gente, del mismo modo que la física fundamental lo es para una concepción reduccionista del mundo. Richard Dawkins opina que «Si criaturas superiores del espacio exterior visitaran la Tierra, la primera pregunta que harían, en orden a evaluar el nivel de nuestra civilización, sería: "¿Han descubierto ya la evolución?"» Muchos

filósofos se han mostrado de acuerdo con René Descartes en que la epistemología es la base de todo conocimiento ulterior y algo parecido al *Cogito, ergo sum* cartesiano es nuestra explicación más básica. Muchos expertos en informática han quedado tan impresionados por las conexiones, recientemente descubiertas, entre física y cálculo, que han concluido que el universo *es* un ordenador que ejecuta programas en forma de leyes físicas. Pero todas estas opiniones acerca de la verdadera estructura de la realidad son limitadas e incluso engañosas. Desde un punto de vista objetivo, la nueva síntesis tiene carácter propio, muy distinto del de cualquiera de las cuatro vías que unifica.

He señalado, por ejemplo, que las teorías fundamentales de las cuatro vías han sido tachadas, en parte con razón, de «inocentes», «estrechas», «frías», etcétera. Así pues, y desde el punto de vista de un físico reduccionista como Stephen Hawking, la especie humana es tan sólo una «escoria química» astrofísicamente insignificante. Steven Weinberg opina que «Cuanto más comprensible parece el universo, más carente de sentido parece también. Pero si no hay alegría en los frutos de la investigación, hay, al menos, consuelo en el propio proceso de investigar» (*The First Three Minutes*, página 154). Pero cualquiera que no esté involucrado en la física fundamental no podrá menos que preguntarse por qué.

Por lo que se refiere al cálculo, el informático Tommasso Toffoli ha señalado que «En ningún caso somos nosotros quienes realizan los cálculos; simplemente, nos montamos por un tiempo en el gran Cálculo que está en marcha». Para él, ello no constituye una expresión de desesperanza, sino todo lo contrario. Pero los críticos de la concepción informática del mundo no desean verse meramente como un programa que alguien ejecuta en un ordenador. La teoría de la evolución, interpretada en sentido restrictivo, nos considera meros «vehículos» para la replicación de nuestros genes o memes, y rehúsa enfrentarse al problema de por qué ha tendido la evolución a crear una complejidad de adaptación siempre creciente, o al del papel que tiene dicha complejidad en el más amplio esquema de las cosas. De modo semejante, la crítica criptoinductivista de la epistemología popperiana se basa en que, mientras establece las condiciones para el desarrollo del conocimiento científico, parece no poder explicar *por qué* crece éste, por qué crea las teorías que utilizamos.

Como he explicado, la defensa consiste, en todos los casos,

en aducir explicaciones tomadas de alguna de las otras vías. No somos *únicamente* «escoria química» puesto que, por ejemplo, el comportamiento general de nuestro planeta, nuestra estrella y nuestra galaxia depende de una magnitud física emergente, pero fundamental: el *conocimiento* que hay en dicha escoria. La creación de conocimiento útil por la ciencia, y de adaptaciones por la evolución, debe ser entendida como la emergencia de la autosemejanza prescrita por un principio de la física, el principio de Turing. Y así sucesivamente.

El problema, pues, de tomar cualquiera de las cuatro teorías fundamentales individualmente como base de una determinada concepción del mundo es que todas son, en un sentido amplio, reduccionistas. Es decir, tienen una estructura explicativa monolítica en la que todo se desprende de algunas ideas extremadamente profundas, lo que deja sin explicar aspectos enteros de la materia correspondiente. En cambio, la estructura explicativa que proporcionan *en conjunto* para la estructura de la realidad no es jerárquica: cada una de ellas contiene principios que resultan «emergentes» desde la perspectiva de las otras tres, pero que, sin embargo, ayudan a explicarlas.

Tres de ellas parecen excluir a los seres y los valores humanos del nivel fundamental de explicación. La cuarta —la epistemología— hace hincapié en el conocimiento, pero no ofrece ninguna razón para que podamos considerarla relevante fuera del contexto de la psicología de nuestra especie. El conocimiento parece un concepto limitado hasta que lo consideramos desde la perspectiva del multiverso. Pero, si el conocimiento tiene una transcendencia fundamental, debemos preguntarnos qué papel parecería natural que desempeñaran en la estructura unificada de la realidad unos seres generadores de conocimiento como nosotros. Esta cuestión ha sido explorada por el cosmólogo Frank Tipler. Su respuesta —la *teoría del punto omega*— constituye un excelente ejemplo de teoría que trata, en el mismo sentido que este libro, de la estructura de la realidad como un todo. No está enmarcada en ninguna de las cuatro vías en concreto, sino que pertenece irreductiblemente a todas ellas. Por desgracia, el propio Tipler, en su libro *La física de la inmortalidad* , hace tan exageradas alabanzas de su teoría que ha provocado que muchos científicos la rechacen de antemano, lo que les hace perderse su valiosa idea central, que paso a explicar.

En mi opinión, el modo más fácil de comprender la teoría del

punto omega es partir del principio de Turing. Es posible un generador universal de realidad virtual. Dicha máquina puede representar cualquier entorno físicamente posible, así como determinadas entidades abstractas hipotéticas, hasta cualquier nivel de exactitud deseado. En consecuencia, su ordenador necesita de una memoria adicional potencialmente ilimitada y puede ejecutar un número ilimitado de pasos. Esto era una cuestión trivial para la teoría clásica de la calculabilidad, a condición de considerar un ordenador universal puramente abstracto. Turing postuló, simplemente, una cinta de memoria de longitud infinita (dotada, según él, de propiedades autoevidentes), un procesador de una exactitud tan perfecta que no requiriese ni energía ni mantenimiento, y disponer de tiempo ilimitado para realizar los cálculos. Hacer más realista a ese modelo previendo un mantenimiento periódico no parece presentar, en principio, mayores problemas, pero las otras tres exigencias —capacidad de memoria, suministro de energía y tiempo de cálculo ilimitados— resultan problemáticos a la luz de la teoría cosmológica actual. Según algunos modelos cosmológicos, el universo volverá a condensarse en un Big Crunch tras un tiempo finito y es, asimismo, espacialmente finito. Tiene la geometría de una «triesfera» (la analogía tridimensional de la superficie en dos dimensiones de una esfera). Semejante cosmología establece un límite finito tanto para la cantidad de memoria que es capaz de acumular la máquina como para el número de pasos que puede ejecutar antes de que se acabe el universo. Ello haría que el ordenador universal fuese físicamente imposible, de modo que el principio de Turing se vería infringido. En otros modelos cosmológicos el universo se expande de manera ilimitada y es espacialmente infinito, lo que parecería facilitar una fuente sin límites de materia para la memoria adicional. Por desgracia, en tales modelos la densidad de la energía disponible para alimentar al ordenador disminuiría con la expansión del universo y debería ser captada cada vez más lejos. Puesto que la física impone como velocidad límite la de la luz, el acceso a la memoria del ordenador debería ralentizarse, y, de nuevo, el efecto neto final sería que tan sólo podría realizarse un número finito de pasos calculatorios.

El descubrimiento clave de la teoría del punto omega consiste en una clase de modelos cosmológicos en que, si bien el universo es finito tanto en espacio como en tiempo, la capacidad de memoria, el número de pasos calculatorios y el suministro efectivo

de energía son ilimitados. Esta imposibilidad aparente resulta posible dada la extrema violencia de los momentos finales del colapso del Big Crunch. Las singularidades en el espacio-tiempo, como el Big Bang y el Big Crunch, no son precisamente lugares tranquilos, pero este último es, con mucho, el peor. En él la forma del universo cambiaría de una triesfera a la analogía tridimensional de la superficie de un elipsoide. El grado de deformación se incrementaría y luego menguaría para volver a crecer después con mayor rapidez en relación a un eje distinto. Tanto la amplitud como la frecuencia de estas oscilaciones aumentarían sin límite al aproximarse el momento final, de modo que habría un número, literalmente, infinito de oscilaciones a pesar de ocurrir dicho final dentro de un tiempo finito. La materia, tal como la conocemos, no sobreviviría. Toda materia, incluyendo los átomos, quedaría destrozada por las tensiones provocadas por las fuerzas gravitacionales generadas por el espacio-tiempo deformado. Sin embargo, estas fuerzas proporcionarían también una fuente ilimitada de energía, que podría, en principio, ser usada para alimentar un ordenador. ¿Cómo podría éste existir en semejantes condiciones? El único «material» que quedaría para poderlo construir serían las partículas elementales y la propia gravedad, es de suponer que en algunos estados cuánticos muy exóticos cuya existencia, al no disponer aún de una adecuada teoría cuántica de la gravedad, no estamos en condiciones de negar o confirmar. (Observarlos de manera experimental queda, obviamente, descartado.) En caso de darse los estados adecuados en partículas y campo gravitatorio, éstos proporcionarían también una ilimitada capacidad de memoria y el universo se encogería tan deprisa que, antes de la llegada del final, sería factible un número ilimitado de accesos a la memoria dentro de un tiempo finito. El punto final del colapso gravitatorio, el Big Crunch de esta cosmología, es lo que Tipler denomina el punto omega.

Ahora bien, el principio de Turing implica que no hay un límite máximo para el número de pasos calculatorios físicamente posibles. Así pues, dado que la clase de cosmología del punto omega es (en asunciones verosímiles) la única en que podría ocurrir un número infinito de pasos calculatorios, podemos inferir que nuestro espacio-tiempo actual debe pertenecer a la clase del punto omega. Puesto que todo cálculo cesaría tan pronto como no hubiera variables capaces de transportar información, podemos inferir que las variables físicas necesarias (quizás

cuantogravitatorias) están presentes hasta que ocurre el punto omega.

Un escéptico podría objetar que esta clase de razonamiento exige una masiva e injustificada extrapolación. Tenemos experiencia de ordenadores «universales» sólo en un entorno muy favorable, que no se parece en nada a los estadios finales del universo. Y esa experiencia se basa únicamente en la ejecución de un número finito de pasos calculatorios para la que se utiliza tan sólo una cantidad finita de memoria. ¿Cómo puede ser válido extrapolar desde estas cantidades finitas hasta el infinito? En otras palabras, ¿cómo podemos saber que el principio de Turing en su versión completa es cierto? ¿Qué evidencia hay de que la realidad respalde una universalidad más que *aproximada*?

Semejante escéptico sería, por supuesto, un inductivista. Más aún, la suya sería exactamente la manera de pensar que (como argumenté en el capítulo anterior) nos impide entender las mejores teorías actuales y avanzar a partir de ellas. Que algo sea o no una «extrapolación» dependerá de la *teoría* de la que partamos. Si partimos de algún concepto vago pero limitado, de lo que es «normal» para las posibilidades de la calculabilidad, un concepto uniformizado por las mejores explicaciones disponibles en la materia, consideraremos que es una «extrapolación injustificada» *cualquier* aplicación de la teoría fuera de las circunstancias habituales. Pero si partimos de explicaciones basadas en la mejor teoría fundamental disponible, lo que nos parecerá una extrapolación injustificada será la propia idea de que se pueda mantener alguna nebulosa «normalidad» en situaciones extremas. Para comprender las mejores teorías actuales, debemos tomarlas en serio como explicaciones de la realidad y no contemplarlas como meros sumarios de las observaciones existentes. El principio de Turing es la mejor teoría actual sobre los fundamentos de la calculabilidad. Por supuesto, conocemos tan sólo un número finito de situaciones que lo confirman, pero ello es igualmente cierto para todas las teorías de la ciencia. Ahí reside —y siempre residirá— la posibilidad lógica de que la universalidad sea de aplicación tan sólo de modo aproximado. Sea como fuere, no existe ninguna teoría rival de la calculabilidad que lo afirme, y ello por una buena razón: un «principio de universalidad aproximada» carecería de todo poder explicativo. Si, por ejemplo, deseamos comprender por qué el mundo *parece* comprensible, la explicación podría ser que el mundo *es* comprensible. Semejante

explicación puede encajar —y encaja, de hecho— con otras explicaciones en otros campos. La teoría de que el mundo es comprensible *a medias*, sin embargo, no explica nada y no podría encajar de ningún modo con otras explicaciones en otros campos, a menos que *éstas* explicaran *aquélla.* Se limita a reformular el problema introduciendo, además, una constante inexplicada: una mitad. En resumen, lo que justifica asumir que el principio de Turing completo se mantiene al final del universo es que cualquier otra asunción estropea las buenas explicaciones sobre lo que sucede aquí y ahora.

Ahora bien, sucede que la clase de oscilaciones en el espacio capaces de originar un punto omega son altamente inestables (al modo del caos clásico), así como violentas. Ambas características se incrementan de modo ilimitado a medida que se acerca dicho punto. Una pequeña desviación de la forma correcta sería magnificada con la suficiente rapidez para que quedara comprometida la continuidad del cálculo, de modo que, después de todo, el Big Crunch ocurriría tras tan sólo un número finito de pasos calculatorios. Por consiguiente, para satisfacer el principio de Turing y alcanzar un punto omega, el universo debería ser «reconducido» continuamente a las trayectorias correctas. Tipler ha demostrado, en principio, que esto se podría conseguir manipulando el campo gravitatorio a través de todo el espacio. Es presumible (necesitaríamos de nuevo una teoría cuántica de la gravedad para confirmarlo) que la tecnología utilizada para la estabilización de mecanismos y el almacenamiento de información debiera ser mejorada sin cesar —de hecho, mejorada un número infinito de veces—, a medida que la densidad y las tensiones aumentaran de modo ilimitado. Ello exigiría la creación continua de conocimiento, lo cual, como nos dice la epistemología popperiana, exige, a su vez, la presencia de la crítica racional y, por consiguiente, de entidades inteligentes. Hemos inferido pues, simplemente a partir del principio de Turing y algunas otras asunciones justificables de manera independiente, que la inteligencia sobrevivirá y el conocimiento continuará creciendo hasta el fin del universo.

Los procedimientos de estabilización, y los correspondientes procesos de creación de conocimiento, deberán ser crecientemente rápidos hasta que, en el frenesí final, ocurra un número infinito de ambos en un tiempo finito. No se conoce ninguna razón por la que no debieran estar disponibles los recursos físicos necesarios para

ello, pero nos podríamos preguntar por qué razón se tomarían tantas molestias los habitantes de ese universo. ¿Por qué tendrían que preocuparse de continuar conduciendo cuidadosamente las oscilaciones gravitatorias, digamos, en el último segundo de su existencia? Cuando a uno le queda tan sólo un segundo de vida, ¿por qué no aprovecharlo para relajarse por fin? Por supuesto, ésta es una interpretación de lo más errónea de la situación. Difícilmente lo podría ser más, ya que las mentes de esas personas estarán funcionando como programas de ordenador en ordenadores cuya velocidad aumentará sin límite. Sus pensamientos serán, como los nuestros, representaciones mediante realidad virtual ejecutadas por dichos ordenadores. Es cierto que, transcurrido ese segundo final, todo el sofisticado mecanismo quedará destruido, pero sabemos ya que la duración subjetiva de una experiencia en realidad virtual no está determinada por el tiempo real transcurrido, sino por la cantidad de cálculos realizados en ese tiempo. En un número infinito de pasos calculatorios hay tiempo para un número infinito de pensamientos, tiempo de sobras para que los pensadores se sitúen en cualquier entorno de realidad virtual que deseen y lo experimenten durante tanto tiempo como quieran. Si se cansan de él, podrán cambiar a otro, o a tantos otros como decidan diseñar. Intrínsecamente, no se encontrarán en los momentos finales de su existencia, sino en el inicio de ésta. No tendrán prisa porque, desde un punto de vista subjetivo, vivirán para siempre. Cuando sólo les quede un segundo, o un microsegundo, dispondrán aún de «todo el tiempo del mundo» para hacer más, experimentar más, crear más —infinitamente más— que nadie antes en el multiverso. No les faltarán, pues, alicientes para dedicar su atención a administrar sus recursos. Y, al hacerlo, no harán más que preparar su propio futuro, un futuro abierto e infinito sobre el cual tendrán pleno control y en el cual, en cualquier momento particular, estarán tan sólo embarcando.

Podemos esperar que en el punto omega la inteligencia esté constituida por nuestros descendientes. Es decir, por nuestros descendientes *intelectuales*, puesto que nuestras formas físicas actuales no podrían sobrevivir en las proximidades del punto omega. En algún momento, los seres humanos deberán transferir los programas de cálculo de sus mentes a algún soporte más resistente. De hecho, esto debería realizarse un número infinito de veces.

La mecánica de «conducir» el universo hacia el punto omega requiere emprender acciones en todo el espacio. De ello se desprende que la inteligencia deberá extenderse a tiempo por todo el universo para realizar los primeros ajustes necesarios. Éste es uno de los retos que Tipler ha mostrado que deberemos afrontar, y, además, ha demostrado que hacerles frente es físicamente posible, de acuerdo con nuestros conocimientos actuales. El primero de dichos retos ocurrirá (como he señalado en el capítulo 8) dentro de unos cinco mil millones de años, cuando el Sol, si se le deja a su aire, se convertirá en una gigante roja que nos destruirá. Deberemos aprender a controlarlo o alejarnos de él antes de que ello suceda. Será necesario después colonizar nuestra galaxia, más tarde el grupo de galaxias del que ésta forma parte y, finalmente, todo el universo. Deberemos hacer frente a cada uno de estos retos en el momento adecuado; no deberemos adelantarnos a los acontecimientos, para no consumir prematuramente los recursos disponibles sin haber desarrollado antes la tecnología correspondiente al próximo nivel.

Digo que «deberemos» hacer todo eso, pero sólo en el sentido en que asumo que seremos nosotros los antepasados de la inteligencia que exista en el punto omega. No estamos obligados a desempeñar ese papel si no lo deseamos. Si decidimos no hacerlo, y el principio de Turing resulta ser cierto, podemos estar seguros de que otros (presumiblemente alguna inteligencia extraterrestre) lo harán.

Mientras tanto, en universos paralelos, nuestras contrapartidas se enfrentan a las mismas opciones. ¿Triunfarán? O, en otras palabras: ¿es *necesario* que alguien consiga crear un punto omega en algún universo? Ello dependerá de lo acertado que sea el principio de Turing. Nos dice que un ordenador universal es físicamente posible, y que «posible», aquí, significa «real en este o algún otro universo». ¿Requiere el principio la construcción de un ordenador universal en todos los universos, sólo en algunos, o quizás en la «mayoría»? No comprendemos aún ese principio lo suficientemente bien para poderlo decidir. Algunos principios de la física, como el de la conservación de la energía, son de aplicación tan sólo en un grupo de universos y pueden ser infringidos en determinadas circunstancias en universos individuales. Otros, como el principio de la conservación de la carga eléctrica, se mantienen estrictamente en todo el multiverso. Las dos formulaciones más simples del principio de Turing serían,

pues: 1) hay un ordenador universal en *todos* los universos; o 2) hay un ordenador universal en, *al menos, algunos* universos. La versión «todos los universos» parece demasiado fuerte para expresar la idea intuitiva de que semejante ordenador sea físicamente *posible.* «Al menos, algunos universos», por otro lado, parece demasiado débil, ya que es evidente que si la universalidad es de aplicación en muy pocos universos pierde por completo su poder explicativo. Pero la versión «la mayoría de los universos» requeriría que el principio especificara un determinado porcentaje, por ejemplo, el 85 por ciento, lo que parece muy poco plausible. (No existen constantes «naturales» en física, si se exceptúan el cero, el uno y el infinito.) Por consiguiente, Tipler opta por «todos los universos», y debo admitir que, dado nuestro conocimiento actual, me parece la elección más lógica.

Esto es todo lo que tiene que decir la teoría del punto omega, o mejor dicho, el componente científico que estoy defendiendo. Podemos llegar a la misma conclusión desde varios puntos de origen distintos en tres de las cuatro vías. Uno de ellos es el principio epistemológico de que *la realidad es comprensible.* Dicho principio es también justificable de modo independiente, en la medida en que constituye la base de la epistemología popperiana. Sus formulaciones existentes, sin embargo, son demasiado vagas para que se puedan extraer de ellas conclusiones categóricas sobre, por ejemplo, la infinitud de las representaciones físicas del conocimiento. Por ello prefiero no postularlo directamente, sino inferirlo a partir del principio de Turing. (He aquí otro ejemplo de la mayor capacidad explicativa disponible al considerar que las cuatro vías constituyen un conjunto fundamental.) El propio Tipler confía o bien en el postulado de que la vida continuará para siempre o bien en el de que lo hará el procesamiento de la información. Desde nuestra perspectiva actual, ninguno de los dos parece fundamental. La ventaja del principio de Turing es que es contemplado ya, por razones muy independientes de la cosmología, como un principio fundamental de la naturaleza, aunque no siempre en su versión más completa, por más que ya he demostrado que ésta resulta indispensable para integrarlo en la física.

Tipler señala que la ciencia de la cosmología ha tendido a estudiar más bien el *pasado* (de hecho, sobre todo, el pasado lejano) del espacio-tiempo, mientras que la mayor parte del espacio-tiempo se encuentra en el futuro de la época actual. Hoy

la cosmología considera la cuestión de si el universo se condensará o no, pero aparte de eso, ha habido muy poca investigación teorética sobre la mayor parte del espacio-tiempo. Lo que conduzca al Big Crunch, en particular, ha merecido mucho menos estudio que las consecuencias del Big Bang. Tipler considera que la teoría del punto omega llena este vacío. Creo que merece convertirse en la teoría dominante sobre el futuro del espacio-tiempo en tanto no sea refutada experimentalmente (o de otro modo). (La refutación experimental es posible, ya que la existencia de un punto omega en el futuro implica determinadas restricciones al estado actual del universo.)

Una vez establecido el escenario para el punto omega, Tipler formula algunas asunciones adicionales —unas plausibles, otras no tanto— que le capacitan para facilitar más detalles de la historia futura. La cuasirreligiosa interpretación que hace de esa historia futura, y su incapacidad para distinguir dicha interpretación de la teoría científica en que se basa, han impedido que esta última sea considerada seriamente. Tipler señala que, para cuando llegue la hora del punto omega, habrá sido generada una cantidad infinita de conocimiento. Asume acto seguido que las inteligencias que existan en ese futuro lejano querrán, como nosotros (o quizás necesitarán) adquirir más conocimiento que el estrictamente necesario para su supervivencia. De hecho, tendrán el potencial para adquirir todo el conocimiento físicamente conocible, y Tipler asume que harán realidad dicho potencial.

Así pues, el punto omega será, en cierto sentido, *omnisciente.*

Pero sólo en cierto sentido. Al atribuir propiedades como la omnisciencia, o incluso la existencia física, al punto omega, Tipler echa mano de un práctico recurso lingüístico bastante corriente en física matemática, pero que puede resultar engañoso si se toma en sentido literal. Consiste en identificar el punto límite de una secuencia con la propia secuencia. Así, cuando dice que el punto omega «sabe» que existe X, significa que X es conocido por alguna entidad finita antes del momento del punto omega y, subsiguientemente, nunca es olvidado. Lo que *no* significa es que exista, en sentido literal, una entidad capaz de conocer en el punto final del colapso gravitatorio, puesto que allí no podrá haber ninguna entidad física. Así, en el sentido más literal, el punto omega no sabe nada, y tan sólo puede decirse que «existe», porque algunas de nuestras explicaciones de la estructura de la

realidad se refieren a las propiedades que limitarán a los sucesos físicos en el futuro lejano.

Tipler emplea el término teológico «omnisciente» por una razón que pronto resultará clara, pero permítaseme señalar de momento que en esa utilización el término no tiene su plena connotación tradicional. El punto omega no lo sabrá *todo.* La abrumadora mayoría de las verdades abstractas, tales como las relativas a los entornos cantgotu y otras semejantes, le resultarán tan inaccesibles como a nosotros.

Ahora bien, dado que la totalidad del espacio estará ocupada por el ordenador inteligente, éste será *omnipresente* (si bien sólo a partir de cierto momento). Puesto que estará constantemente reconstruyéndose y guiando el colapso gravitacional, puede considerarse que tendrá el control sobre todo lo que suceda en el universo material (o en el multiverso, si el punto omega sucede en todos los universos). Por consiguiente, dice Tipler, será también *omnipotente.* Pero, una vez más, esta omnipotencia no será absoluta. Bien al contrario, estará estrictamente limitada por la materia y la energía disponibles, y sujeta a las leyes de la física.

Puesto que las inteligencias del ordenador serán pensadores creativos, deben ser clasificados como «personas». Cualquier otra clasificación, argumenta acertadamente Tipler, resultaría racista. Y, por ello, afirma que en el límite del punto omega existirá una sociedad de personas omnisciente, omnipresente y omnipotente. Tipler identifica a esta sociedad con Dios.

He mencionado algunos de los aspectos en que el «Dios» de Tipler difiere del Dios o dioses en que creen la mayoría de las personas religiosas. Hay otras diferencias. Por ejemplo, las personas próximas al punto omega no podrían, aunque quisieran, hablarnos, comunicarnos sus deseos o hacer milagros (hoy). No crearon el universo ni inventaron las leyes de la física, y no las podrían infringir aunque quisieran. Pueden escuchar nuestras súplicas actuales (quizás detectando señales muy débiles), pero no pueden responder a ellas. Son (lo podemos inferir de la epistemología popperiana) contrarias a la fe religiosa y no desean ser adoradas. Y así sucesivamente. Pero Tipler no se detiene aquí y argumenta que la mayor parte de las propiedades fundamentales del Dios de las diversas confesiones religiosas derivadas de la tradición judeocristiana lo son también del punto omega. Supongo que la mayoría de las personas religiosas discreparán de Tipler en lo que concierne a las características básicas de sus respectivas

confesiones.

En particular, Tipler afirma que una tecnología lo suficientemente avanzada será capaz de resucitar a los muertos. Podrá hacerlo de diferentes maneras, de las cuales la siguiente es quizás la más sencilla. Una vez que se disponga de la suficiente capacidad calculatoria (recordemos que llegará un momento en que estará disponible en cualquier cantidad que se desee), se podrá representar mediante realidad virtual todo el universo —de hecho, todo el multiverso—, a partir del Big Bang, con cualquier grado de fidelidad deseado. Si no se conoce el estado inicial con la suficiente exactitud, siempre se podrá ensayar con un muestrario lo más aproximado que se pueda de todos los estados iniciales posibles representados de manera simultánea. La reproducción podría tener que hacer una pausa, por razones de complejidad, si la época representada se acercara demasiado al tiempo real en el que tuviera lugar la representación. Pero pronto podría reanudar la ejecución, a medida que se fuera incorporando mayor capacidad de cálculo. Para los ordenadores del punto omega, nada es intratable. Sólo existe lo «calculable» y lo «no calculable», y la representación de entornos físicos reales queda definitivamente dentro de la primera categoría. En el transcurso de la representación aparecerían el planeta Tierra y múltiples variantes suyas. La vida y, en su momento, los seres humanos, evolucionarían. Todos los seres humanos que hayan vivido en algún momento en cualquier lugar del multiverso (es decir, todos aquellos cuya existencia haya sido físicamente posible) aparecerían a su debido tiempo en tan vasta representación, al igual que toda inteligencia artificial o extraterrestre que haya existido jamás. El programa que ejecutara la representación podría localizar a esos seres inteligentes y, si lo deseara, situarlos en un entorno virtual mejor, en el que quizás no volvieran a morir y vieran satisfechos todos sus deseos (o, al menos, todos los que un determinado nivel de cálculo, inimaginablemente elevado, pudiera satisfacer). ¿Por qué tendría que hacer una cosa así? Una razón podría ser de índole moral: según los estándares del futuro lejano, el entorno en el que vivimos hoy día es muy duro y, en consecuencia, sufrimos de un modo atroz. Podría ser considerado poco ético no rescatar a estos infelices para proporcionarles la oportunidad de una vida mejor. Sin embargo, podría resultar contraproducente ponerlos en contacto con la cultura contemporánea en el momento inmediato a su resurrección: se

sentirían confundidos, desplazados y humillados. Por consiguiente, dice Tipler, sería deseable que resucitaran en un entorno básicamente familiar, pero del que se hubiera retirado cualquier elemento desagradable y al que se hubieran incorporado toda clase de situaciones placenteras. En otras palabras, el cielo.

Tipler continúa de idéntica manera y reestructura muchos otros aspectos del escenario religioso tradicional, que redefine como entidades o procesos físicos susceptibles de existir cerca del punto omega. Dejemos de lado la cuestión de si esas versiones reestructuradas son fieles o no a sus análogas religiosas. La historia de lo que harán o dejarán de hacer esas inteligencias del lejano futuro se basa en una cadena de suposiciones. Incluso en el caso de conceder que fueran individualmente plausibles, las conclusiones generales sacadas de ellas no dejan de ser más que especulación erudita. Puede ser interesante elaborar tales suposiciones, pero sin perder de vista la importancia de saberlas diferenciar de la argumentación para justificar la propia existencia del punto omega, así como de la teoría de sus propiedades físicas y epistemológicas, ya que *esos* argumentos sólo asumen que la estructura de la realidad se ajusta a las mejores teorías de que disponemos, asunción que puede ser justificada de modo independiente.

Como una advertencia sobre la poca fiabilidad de la especulación, incluso erudita, visitemos de nuevo a nuestro maestro constructor del capítulo 1, con su conocimiento precientífico de la arquitectura y la ingeniería. Nos separa de él un lapso cultural tan grande, que le resultaría, sin duda, en extremo difícil concebir una imagen aceptable de nuestra civilización. Sin embargo, somos prácticamente contemporáneos en comparación con la enorme distancia que nos separa del primer momento posible para la resurrección de Tipler. Supongamos ahora que nuestro constructor especula acerca del futuro lejano de la industria de la construcción y, por algún golpe de suerte extraordinario, elabora una evaluación perfectamente ajustada de la tecnología actual. Sabrá entonces, entre otras cosas, que somos capaces de construir estructuras mucho más grandiosas e impresionantes que las mayores catedrales de su época. Podríamos construir una catedral de un kilómetro y medio de altura, si lo deseáramos, y sería posible hacerlo utilizando una proporción mucho menor de nuestra riqueza, y mucho menos tiempo y esfuerzo humano, que los que él necesitaba para

construir una catedral de lo más modesto. Podría haberse sentido, pues, seguro al afirmar que, hacia el año 2000, existirían catedrales de un kilómetro y medio de altura. Sin embargo, se habría equivocado totalmente ya que, si bien disponemos de la tecnología necesaria para construir tales estructuras, no tenemos motivos para hacerlo. De hecho, ahora parece improbable que llegue a construirse jamás una catedral así. Aun suponiendo que hubiese acertado en lo relativo al estado de nuestra tecnología, nuestro casi contemporáneo se habría equivocado de medio a medio acerca de nuestras motivaciones, y ello porque algunas de sus más incuestionadas asunciones sobre lo que incitaba a obrar a los seres humanos habrían quedado obsoletas transcurridos tan sólo algunos siglos.

De modo semejante, nos podría parecer natural que las inteligencias del punto omega, por razones de investigación histórica o arqueológica, o por compasión, deber moral o mero capricho, crearan, llegado el caso, representaciones de nosotros en realidad virtual, y que, una vez terminado su experimento, nos donaran los insignificantes recursos calculatorios que necesitaríamos para vivir para siempre en el «cielo». (Personalmente, preferiría que me permitiesen adaptarme de modo gradual a su cultura.) Pero no podemos saber lo que querrán. De hecho, ningún intento de profetizar acontecimientos futuros a gran escala referido a asuntos humanos (o sobrehumanos) puede proporcionar resultados fiables. Como Popper ha señalado, el curso futuro de los asuntos humanos depende del futuro desarrollo del conocimiento. No podemos predecir qué conocimiento específico será creado en el futuro, ya que, de poderlo hacer, por definición, lo poseeríamos.

No sólo el conocimiento científico conforma las preferencias de las personas y determina cómo deciden comportarse. Existen también, por ejemplo, los criterios morales, que asignan a las posibles acciones atributos tales como «bueno» o «malo». Dichos valores han resultado especialmente difíciles de acomodar a la visión científica. Parecen constituir por sí mismos una estructura explicativa cerrada, desconectada de la del mundo físico. Como señaló David Hume, es imposible derivar lógicamente un «debe ser» de un «es». Sin embargo, utilizamos tales valores tanto para explicar nuestras acciones físicas como para determinarlas.

El pariente pobre de la moralidad es la *utilidad.* Puesto que parece mucho más fácil entender qué es objetivamente útil o inútil

que decidir qué es bueno o malo, han proliferado los intentos de definir la moralidad en términos de diversas formas de utilidad. Hay, por ejemplo, una moralidad evolutiva que señala que muchas de las formas de comportamiento que explicamos en términos morales, como no matar o no hacer trampas en los tratos con otras personas, tienen sus análogos en el comportamiento de los animales. Una rama de la teoría evolutiva, la *sociobiología*, ha conseguido algunos éxitos en la explicación del comportamiento animal. Mucha gente se ha sentido tentada a concluir que las explicaciones morales para las opciones humanas son simples adornos, que la moralidad carece de base objetiva y que «bueno» y «malo» son simples etiquetas que aplicamos a los impulsos innatos que nos llevan a comportamientos de un modo u otro. Otra versión de esta explicación sustituye los genes por memes y afirma que la terminología moral no es más que un simple decorado para el condicionamiento social. Sin embargo, ninguna de esas explicaciones se ajusta a los hechos. Por una parte, *no* tendemos a explicar el comportamiento innato —los ataques epilépticos, por ejemplo— en términos de elección moral; sabemos distinguir entre acto voluntario e involuntario, y buscamos explicaciones tan sólo para la primera categoría. Por otra, no hay ni un solo comportamiento innato —evitar el dolor, tener relaciones sexuales, comer o cualquier otro— al que los seres humanos no hayan renunciado por razones morales en determinadas circunstancias. Lo mismo es de aplicación, y de manera aún más común, en relación con el comportamiento condicionado socialmente. De hecho, el rechazo de los comportamientos innatos, así como de los condicionados socialmente, es, en sí, una característica del modo de obrar humano, al igual que lo es explicar esas rebeliones en términos morales. Ninguno de esos comportamientos tiene analogías entre los animales, y en ninguno de esos casos pueden las explicaciones morales ser reinterpretadas en términos genéticos o meméticos. Se trata de un error fundamental, común a todas las teorías de esa clase. ¿Podría existir un gen que nos permitiese rechazar la conducta de los demás genes cuando así lo decidiéramos? ¿O un condicionamiento social que fomentase la rebelión? Quizás, pero ello dejaría intacto el problema de *cómo escogimos hacer lo contrario de lo que se suponía que debíamos realizar* y de qué queremos expresar al explicar nuestra rebelión manifestando que escogimos, simplemente, lo correcto, y que lo que nuestros genes

o nuestra sociedad nos prescribían era incorrecto.

Estas teorías genéticas pueden ser contempladas como un caso especial de una estratagema más amplia, consistente en negar que los juicios morales sean significativos alegando que no escogemos realmente nuestras acciones y que el libre albedrío es pura ilusión, incompatible con la física. Pero, en realidad, y como vimos en el capítulo anterior, el libre albedrío *es* compatible con la física y encaja de modo muy natural en la estructura de la realidad que he descrito.

El *utilitarismo* fue un intento anterior de integrar las explicaciones morales en el contexto científico mediante la «utilidad». Ésta se asimilaba a la felicidad humana. Las elecciones morales debían basarse en el cálculo de qué acción produciría más felicidad, ya fuera para el individuo o (y ahí la teoría se hacía más vaga) para el «mayor número» posible de personas. Distintas versiones de la teoría sustituían «felicidad» por «placer» o «preferencia». Considerado un repudio de los precedentes sistemas autoritarios de moralidad, el utilitarismo es intachable. Y, en el sentido de que, simplemente, aboga por rechazar los dogmas y obrar según la teoría «preferida», la que haya superado la crítica racional, cualquier ser racional es utilitarista. Pero en cuanto intento de resolver el problema que estamos analizando, es decir, explicar el significado de los juicios morales, tiene también un defecto fatal: *escogemos nuestras preferencias.* En particular, *cambiamos* de preferencias, y aducimos explicaciones morales para hacerlo. Semejantes explicaciones no pueden ser traducidas a términos utilitaristas. ¿Existe una preferencia maestra oculta que controla los cambios de preferencia? De ser así, sería imposible cambiarla, y el utilitarismo degeneraría en la teoría genética de la moralidad, de la que hemos hablado antes.

¿Cuál es, pues, la relación entre los valores morales y la particular concepción científica del mundo que defiendo en el presente libro? Cuando menos, puedo argumentar que no existe ningún obstáculo fundamental para formular una. El problema de todas las «concepciones científicas del mundo» anteriores era que tenían estructuras explicativas jerárquicas. Del mismo modo que resulta imposible, dentro de semejante estructura, «justificar» que las teorías científicas sean *ciertas*, tampoco es posible justificar que una determinada línea de acción sea *correcta* (ya que entonces, ¿cómo se justificaría que la estructura, en su totalidad, lo sea?). Como he dicho, cada una de las cuatro vías tiene una

estructura explicativa jerárquica propia, pero la estructura total de la realidad no. Explicar, pues, los valores morales como atributos objetivos de los procesos físicos no tiene por qué equivaler a derivarlos de algo, incluso en principio. Al igual que sucede con las entidades matemáticas abstractas, la cuestión será ver con qué contribuyen a la explicación y si la realidad física puede ser comprendida o no sin atribuir también realidad a tales valores.

A este respecto, permítaseme señalar que la «emergencia», en el sentido estándar, es tan sólo uno de los modos en que pueden estar relacionadas las explicaciones de las distintas vías. Hasta aquí, he considerado realmente tan sólo lo que podríamos denominar emergencia *predictiva.* Por ejemplo, creemos que las predicciones de la teoría de la evolución se desprenden lógicamente de las leyes de la física, aunque demostrar la conexión mediante el cálculo pueda resultar intratable. No pensamos, en cambio, que las *explicaciones* de la teoría de la evolución se desprendan de la física. Sin embargo, una estructura explicativa no jerárquica admitiría la posibilidad de la emergencia explicativa. Supongamos, en aras de la argumentación, que un determinado juicio moral pueda ser explicado como bueno en algún estrecho sentido utilitario. Por ejemplo: «Lo deseo, no perjudica a nadie, luego es bueno.» Ahora bien, ese juicio podría ser cuestionado algún día. Podría preguntarme: «¿Debería desearlo?», a: «¿Cómo puedo estar seguro de que no perjudica a nadie?», ya que la cuestión de a quién juzgo que la acción pueda «perjudicar» se basa en presunciones morales. Que permanezca repantigado en mi sillón «perjudica» de modo directo a todos los seres humanos que podrían beneficiarse de que corriera en su ayuda, a todos los ladrones que podrían intentar robarme el sillón si me levantara de él, y así sucesivamente. Para resolver estas cuestiones recurro a teorías morales adicionales, que incluyen nuevas explicaciones de mi situación moral. Cuando alguna de ellas parece satisfactoria, la utilizo de modo provisional para formular juicios de bondad o maldad, pero la explicación, si bien temporalmente satisfactoria para mí, sigue sin ir más allá del nivel utilitario.

Pero supongamos ahora que alguien elabora una teoría general de esas explicaciones en cuanto tales, introduce en ella un concepto de nivel superior, como «derechos humanos», *y* considera que la introducción de ese concepto proporcionará siempre, para una clase dada de problemas, como la que acabo de

describir, una nueva explicación que resolverá el problema en sentido utilitario. Supongamos, incluso, que esa teoría de las explicaciones es, a su vez, una teoría explicativa. Explica, en términos de alguna otra vía, *por qué* es «mejor» (en el sentido utilitario) analizar los problemas en términos de derechos humanos. Por ejemplo, podría explicar sobre bases epistemológicas por qué cabe esperar que el respeto por los derechos humanos promueva el desarrollo del conocimiento, lo cual es en sí una condición previa para la resolución de problemas morales.

Si la explicación parece buena, podría valer la pena adoptar la teoría. Es más, puesto que los cálculos utilitarios son tremendamente difíciles de realizar, mientras que, por lo general, resulta factible analizar una situación en términos de derechos humanos, podría resultar conveniente utilizar un análisis basado en ellos con preferencia a cualquier teoría específica sobre las posibles implicaciones de una determinada acción. Si todo ello fuera cierto, podría ser que el concepto de «derechos humanos» no fuera expresable, incluso en principio, en términos de «felicidad», es decir, que no fuera un concepto utilitario. Podríamos denominarlo un concepto moral. La conexión entre los dos no quedaría establecida por la predicción emergente, sino por la explicación emergente.

No es que abogue por este enfoque concreto; simplemente, expongo la manera en que los valores morales podrían existir de modo objetivo al contribuir a las explicaciones emergentes. Si este enfoque funcionara, explicaría la moralidad como una especie de «utilidad emergente».

De modo parecido, el «valor artístico» *y* los demás conceptos estéticos han resultado siempre difíciles de explicar en términos objetivos. También son explicados, a menudo de modo superficial, como facetas arbitrarias de la cultura o en términos de preferencias innatas. De nuevo, vemos que ello no tiene por qué ser así. Al igual que la moralidad está relacionada con la utilidad, el valor artístico posee una contrapartida menos elevada, pero más objetivamente definible. Se trata del *diseño.* Una vez más, el valor de un diseño es tan sólo comprensible en el contexto de un propósito determinado para el objeto diseñado. Nos podemos encontrar, no obstante, con que resulte factible mejorar un diseño mediante la incorporación de buenos criterios estéticos a los criterios de diseño. Tales criterios estéticos serían incalculables a

partir de los criterios de diseño, ya que una de sus utilidades sería la mejora de éstos. La relación sería, una vez más, de emergencia explicativa, y el valor artístico, o belleza, constituiría una especie de *diseño emergente.*

La excesiva confianza de Tipler al predecir las motivaciones de la gente cuando se acerque el punto omega le lleva a subestimar una importante implicación de su teoría en el papel de la inteligencia en el universo. La inteligencia no sólo estará allí para controlar los acontecimientos a gran escala, sino también para decidir lo que vaya a suceder. Como dijo Popper, nosotros decidiremos cómo se acabará el universo. Y, de hecho, en gran medida, el contenido de los pensamientos inteligentes futuros *es* lo que sucederá, puesto que, al final, la totalidad del espacio y su contenido *serán* el ordenador. El universo consistirá al final, literalmente, en procesos de pensamiento inteligentes. En algún lugar, hacia el lejano final de esos pensamientos materializados, quizás se halle todo el conocimiento físicamente posible, expresado mediante conformaciones físicas.

Las deliberaciones morales y estéticas también estarán expresadas en dichas conformaciones, así como los resultados de tales deliberaciones. En efecto, exista o no un punto omega, dondequiera que haya conocimiento en el multiverso (entendido como complejidad extendida por múltiples universos), debe haber también huellas físicas del razonamiento moral y estético que determinó qué clase de problemas eligió resolver allí la entidad creadora de conocimiento. En particular, antes que cualquier porción de conocimiento factual pueda extenderse con un contenido homogéneo por una serie de universos, los juicios morales y estéticos deben ser ya similares en dichos universos. De ello se desprende que tales juicios contienen también conocimiento objetivo, en el sentido físico y multiversal. Ello justifica la utilización de terminología epistemológica como «problema», «solución», «razonamiento» y «conocimiento» en ética y estética. Así, si éstas han de ser compatibles de algún modo con la concepción del mundo por la que abogo en el presente libro, la belleza y la bondad deben ser tan objetivas como la verdad matemática o científica, y deben ser creadas de modo análogo: mediante la conjetura y la crítica racional.

Keats, pues, estaba en lo cierto al afirmar que «belleza es verdad, verdad es belleza». No pueden ser una misma cosa, pero son una misma *clase* de cosa, creadas de modo similar e

inseparablemente relacionadas. (Se equivocaba, sin embargo, al añadir que «eso es todo cuanto puedes conocer del mundo y todo cuanto necesitas conocer de él».)

Llevado de su entusiasmo (¡en su sentido primigenio de posesión divina!), Tipler omite parte de las enseñanzas popperianas sobre el aspecto que debe presentar el desarrollo del conocimiento. De existir el punto omega y ser como Tipler propone, el universo final consistirá, sin duda, en pensamientos corporeizados de inconcebible sabiduría, creatividad y multiplicidad. Pero pensamiento es resolución de problemas, lo que significa conjeturas rivales, error, crítica, refutación y retroceso. Sin duda, *en el límite* (que nadie experimentará), en el instante mismo en que se acabe el universo, todo lo comprensible habrá podido ser comprendido, pero en cada punto finito anterior el conocimiento de nuestros descendientes estará plagado de errores. Será mayor, más profundo y más extenso de lo que nunca podamos imaginar, pero cometerán también errores a una escala proporcionalmente titánica.

Como nosotros, no conocerán jamás la certeza o la seguridad física, ya que su supervivencia, al igual que la nuestra, dependerá de que creen un flujo continuo de nuevo conocimiento. Si fracasan, aunque sólo sea una vez, llegado el momento de hallar el modo de incrementar su velocidad de cálculo y su capacidad de memoria dentro del tiempo disponible, determinado por una ley física inexorable, el cielo se desplomará sobre sus cabezas y perecerán. Su cultura será, presumiblemente, pacífica y benevolente más allá de nuestros sueños más atrevidos, pero no tranquila. Estará involucrada en la resolución de problemas terribles y lacerada por controversias apasionadas. Por esta razón, no parece probable que pueda ser útil considerarla una «persona». Consistirá más bien en un vasto número de personas que interactuarán en múltiples niveles y de muchas maneras distintas, pero en *desacuerdo.* No hablarán con una sola voz, del mismo modo que no lo hacen hoy día los científicos participantes en un seminario de investigación. Incluso si, por casualidad, llegan a ponerse de acuerdo, sus opiniones serán a menudo erróneas, y estos errores perdurarán durante períodos arbitrariamente largos (desde un punto de vista subjetivo). Por la misma razón, nunca será su cultura *moralmente* homogénea. Nada será sagrado (¡otra diferencia, sin duda, con la religión convencional!), y habrá personas que cuestionarán sin tregua asunciones consideradas por

otras como verdades morales fundamentales. Por supuesto, la moralidad, al ser real, es comprensible mediante los métodos de la razón, y, por consiguiente, toda controversia particular será resuelta. Pero, inmediatamente, la sustituirá por otra mayor, más fundamental y apasionante. Una colección tan discordante —aunque progresista— de comunidades solapadas es muy distinta del Dios en el que creen las personas religiosas, pero será, precisamente (o, más bien, alguna subcultura que forme parte de ella) lo que nos resucite, si Tipler está en lo cierto.

En vista de todas las ideas unificadoras que he expuesto, como el cálculo cuántico, la epistemología evolutiva y las concepciones multiversales del conocimiento, el libre albedrío y el tiempo, me parece evidente que la tendencia actual en nuestra comprensión general de la realidad se ajusta a lo que, siendo niño, yo esperaba que fuera. Nuestro conocimiento es cada vez más amplio y más profundo, y predomina en él, como dije en el capítulo 1, la profundidad. He abogado, sin embargo, por algo más en este libro. He abogado por una determinada concepción unificada del mundo basada en las cuatro vías: la física cuántica del multiverso, la epistemología de Karl Popper, la teoría de Darwin-Dawkins de la evolución y la versión completa de la teoría de la calculabilidad universal de Turing. Me parece que, en el estado actual de nuestro conocimiento científico, ésta es la visión que resulta «natural» adoptar. Se trata de una visión conservadora, que no impone ningún cambio radical en las mejores explicaciones fundamentales de que disponemos. Debería, por lo tanto, constituir la concepción dominante, aquella por oposición a la cual deberían ser juzgadas las innovaciones propuestas. Éste es, precisamente, el papel que reclamo para ella. Lejos de mí cualquier intención de fundar una nueva ortodoxia. Como he dicho, creo que es hora de ponerse en marcha, pero sólo podremos encontrar mejores teorías en nuestro camino si nos tomamos en serio las mejores teorías de que disponemos en la actualidad como explicaciones del mundo.

BIBLIOGRAFÍA

LECTURAS FUNDAMENTALES

Richard Dawkins, *The Selfish Gene*, Oxford University Press, 1976. (Edición revisada 1989.) [Edición en castellano: *El gen egoísta*, Biblioteca Científica Salvat, vol. 5. Salvat, 1993.]

Richard Dawkins, *The Blind Watchmaker*, Norton, 1987; Penguin Books, 1990. [Edición en castellano: *El relojero ciego*, Labor, 1989.]

David Deutsch, «Comment on "The Many Minds Interpretation of Quantum Mechanics", by Michael Lockwood», *British Journal for the Philosophy of Science*, 1996, vol. 47, n° 2, p. 222.

David Deutsch y Michael Lockwood, «The Quantum Physics of Time Travel», *Scientific American*, marzo de 1994, p. 68.

Douglas R. Hofstadter, *Gödel, Escher, Bach, an Eternal Golden Braid*, Harvester, 1979; Vintage Books, 1980. [Edición en castellano: *Gödel, Escher, Bach*, 3.a ed., Tusquets, 1989.]

James P. Hogan, *The Proteus Operation*, Baen Books, 1986; Century Publishing, 1986. (¡Ficción!) [Edición en castellano: *Herederos de las estrellas*, Edhasa, 1986.]

Bryan Magee, *Popper*, Fontana, 1973; Viking Penguin, 1995.

Karl Popper, *Conjectures and Refutations*, Routledge, 1963; Harper Collins, 1995. [Edición en castellano: *Conjeturas y refutaciones*, Paidós Ibérica, 1991.]

Karl Popper, *The Myth of the Framework*, Routledge, 1992. [Edición en castellano: *El porvenir está abierto*, Tusquets, 1992.]

LECTURAS RECOMENDADAS

John Barrow y Frank Tipler, *The Anthropic Cosmological Principle*, Clarendon Press, 1986.

Charles H. Bennett, Gilíes Brassard y Artur K. Ekert, «Quantum Cryptography», *Scientific American*, octubre de 1992.

Jacob Bronowski, *The Ascent of Man*, BBC Publications, 1981; Little Brown, 1976.

Julian Brown, «A Quantum Revolution for Computing», *New Scientist*, 24 de septiembre de 1994.

Paul Davies y Julian Brown, *The Ghost in the Atom*, Cambridge University Press, 1986.

Richard Dawkins, *The Extended Phenotype*, Oxford University Press, 1982.

Daniel C. Dennett, *Darwin s Dangerous Idea: Evolution and the Meanings of Life*, Allen Lane, 1995; Penguin Books, 1996.

Bryce S. DeWitt y Neill Graham (eds.), *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton University Press, 1973.

Artur K. Ekert, «Quantum Keys for Keeping Secrets», *New Scientis*t, 16 de enero de 1993.

Freedom and Rationality: Essays in Honour of John Watkins, Kluwer, 1989.

Ludovico Geymonat, *Galileo Galilei: A Biography and Inquiry into his Philosophy of Science*, McGraw-Hill, 1965. [Edición en castellano: *Galileo Galilei*, Edicions 62, 1986.]

Thomas Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, 1971. [Edición en castellano: *Qué son las revoluciones científicas y otros ensayos*, Paidós Ibérica, 1989.]

Imre Lakatos y Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*, Cambridge University Press, 1979. [Edición en castellano: *Pruebas y refutaciones: la lógica del descubrimiento matemático*, Alianza, 1994.]

Seth Lloyd, «Quantum-mechanical Computers», *Scientific American*, octubre de 1995.

Michael Lockwood, *Mind, Brain and the Quantum*, Basil

Blackwell, 1989.

Michael Lockwood, «The Many Minds Interpretation of Quantum Mechanics», *British Journal for the Philosophy of Science*, vol. 47, n.°2, 1996.

David Miller (ed.), *A Pocket Popper*, Fontana, 1983.

pavid Miller, *Critical Rationalism: A Restatement and Defense*, Open Court, 1994.

Ernst Nagel y James R. Newman, *Godel's Proof* Routledge, 1976.

Anthony O'Hear, *Introduction to the Philosophy of Science*, Oxford University Press, 1991.

Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford University Press, 1989.

Karl Popper, *Objective Knowledge: An Evolutionary Approach*, Clarendon Press, 1972. [Edición en castellano: *Conocimiento objetivo*, Tecnos, 1988.]

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech y Jan Svartvik, *A Comprehensive Grammar of the English Language*, 7.a ed., Longman, 1989.

Dennis Sciama, *The Unity of the Universe*, Faber & Faber, 1967.

Ian Stewart, *Does God Play Dice? The Mathematics of Chaos*, Basil Blackwell, 1989; Penguin Books, 1990. [Edición en castellano: *¿Juega Dios a los dados?*, RBA Ed., 1994.]

L. J. Stockmeyer y A. K. Chandra, «Intrinsically Difficult Problems», *Scientific American*, mayo de 1979.

Frank Tipler, *The Physics of Immortality*, Doubleday, 1995.

Alan Turing, «Computing Machinery and Intelligence», *Mind*, octubre de 1950. (Reimpreso en *The Mind's I*, editado por Douglas Hofstadter y Daniel C. Dennett, Harvester, 1981.)

Steven Weinberg, *Gravitation and Cosmology*, John Wiley, 1972.

Steven Weinberg, *The First Three Minutes*, Basic Books, 1977. [Edición en castellano: *Los tres primeros minutos del universo*, Alianza, 1992.]

Steven Weinberg, *Dreams of a Final Theory*, Vintage, 1993; Random, 1994.

John Archibald Wheeler, *A Journey into Gravity and Spacetime, Scientific American Library*, 1990.

Lewis Wolpert, *The Unnatural Nature of Science*, Faber & Faber, 1992; HUP, 1993.

Benjamin Woolley, *Virtual Worlds*, Basil Blackwell, 1992; Penguin Books, 1993.

Notas

[1]Del inglés *to feel*, «sentir, notar». En adelante, utilizaré los términos *película sensorial, máquina sensorial, tecnología sensorial*, etcétera, para hacer referencia a esta propiedad. (*N. del T.*). <<

[2]*Cantgotu* en inglés, que se pronuncia «cantgotu». (*N. del T*.)<<

[3]En Freedom and Rationality: Essays in Honour of John Watkins, Klurver, Dordrecht—Boston, 1989.<<

[4]En realidad, los teoremas matemáticos tampoco son probados mediante la *pura* argumentación (independiente de la física), como explicaré en el capítulo 10. (*N. del T*.)<<

[5]En realidad, podría seguir siendo universalmente cierta, si otras teorías acerca de la estructura y condiciones del experimento fuesen falsas. (*N. del T*.)<<

[6]En realidad, podría seguir siendo universalmente cierta, si otras teorías acerca de la estructura y condiciones del experimento fuesen falsas. (*N. del T*.)<<